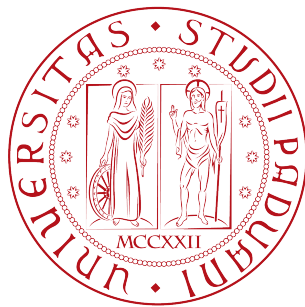


Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in  
Statistica per le Tecnologie e le Scienze



RELAZIONE FINALE  
**LINK PARAMETRICI NEI GLM: TEORIA ED  
APPLICAZIONI**

Relatore Prof.ssa Laura Ventura  
Dipartimento di Scienze Statistiche

Laureando: Anna Montin  
Matricola N. 1145449

Anno Accademico 2018/2019



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Modelli Lineari Generalizzati</b>	<b>3</b>
1.1 Modello di regressione lineare normale . . . . .	3
1.1.1 Funzione di verosimiglianza e stima dei parametri del modello . . . . .	5
1.1.2 Inferenza sul modello . . . . .	6
1.1.3 Coefficiente di determinazione $R^2$ . . . . .	7
1.1.4 Residui e bontà d'adattamento . . . . .	7
1.2 Modello di regressione lineare generalizzato . . . . .	8
1.2.1 Famiglie di dispersione esponenziale . . . . .	8
1.2.2 Media e Varianza . . . . .	9
1.2.3 Componenti di un modello lineare generalizzato . . . .	10
1.2.4 Funzione di legame . . . . .	11
1.2.5 Verosimiglianza e inferenza . . . . .	12
1.2.6 Informazione osservata e attesa . . . . .	14
1.2.7 Minimi quadrati pesati iterati . . . . .	15
1.2.8 Devianza e bontà d'adattamento . . . . .	16
1.2.9 Residui . . . . .	18
1.3 Conclusioni . . . . .	19
<b>2 Funzioni di legame parametriche</b>	<b>21</b>
2.1 GLM con link parametrici . . . . .	22
2.1.1 Ortogonalizzazione dei parametri nei GLM con link parametrici . . . . .	23
2.1.2 Standardizzazione di famiglie di link parametrici nei GLM . . . . .	25
2.2 Famiglie parametriche a uno e due parametri . . . . .	26
2.2.1 Famiglie di trasformazioni di potenza . . . . .	27
2.2.2 Famiglie di trasformazioni per la regressione binaria .	29
2.3 Quando usare un link non canonico . . . . .	33
2.3.1 Test d'ipotesi per GLM con link parametrici . . . . .	33
2.3.2 Procedura per l'identificazione del link corretto . . . .	36

---

2.4	Conclusioni . . . . .	37
<b>3</b>	<b>Stima dei parametri di link</b>	<b>39</b>
3.1	Teoria della stima del parametro di legame . . . . .	39
3.1.1	Un esempio . . . . .	42
3.2	Stima del parametro di legame con LikelihoodAsy . . . . .	43
3.3	Modifica delle code della funzione di legame . . . . .	45
3.3.1	Regressione logistica modificata . . . . .	45
3.3.2	Confronto di due modifiche della funzione logistica . .	48
3.4	Conclusioni . . . . .	51
	<b>Conclusioni</b>	<b>52</b>
	<b>Bibliografia</b>	<b>55</b>
	<b>Appendice</b>	<b>56</b>

# Elenco delle figure

2.2.1	Modifica di entrambe le code per diversi valori di $\psi_1$ e $\psi_2$ con $\eta_0 = 0$ .	27
2.2.2	Modifica della coda destra per diversi valori di $\psi_1$ con $\eta_0 = 0$ .	28
2.2.3	Modifica della coda sinistra per diversi valori di $\psi_2$ con $\eta_0 = 0$ .	28
2.2.4	Modifica simmetrica delle code per diversi valori di $\psi_1$ e $\psi_2$ con la famiglia di trasformazioni di Stukel per dati binari.	32
2.2.5	Media in funzione del predittore lineare per diversi valori di $\psi_1$ e $\psi_2$ con la famiglia di trasformazioni di Stukel per dati binari.	32
3.2.1	Log-verosimiglianza profilo con funzione parametrica di Burr-Prentice al variare del parametro $\psi$ .	44
3.2.2	Devianza per il confronto tra il modello con funzione parametrica di Burr-Prentice con il modello logistico al variare del parametro $\psi$ .	45
3.3.1	Devianza per il modello di regressione logistica per vari valori di $\psi_2$ e con $\eta_0 = 0$ .	48
3.3.2	Modello logistico e modello logistico con coda modificata con $\psi_2 = 0.17$ e $\eta_0 = 0$ .	49
3.3.3	Devianza per il modello di regressione logistica per vari valori di $\psi_1$ , a destra, e di $\psi_2$ a sinistra, con $\eta_0 = 0$ .	50
3.3.4	Modello logistico e modelli modificati con $\psi_1 = 1.92$ e $\psi_2 = 0.16$ e con $\eta_0 = 0$ .	51
3.4.1	Curve di livello per la devianza del modello di regressione con funzione di legame probit per vari valori di $\psi_1$ e $\psi_2$ con $\eta_0 = 0$ .	54



# Elenco delle tabelle

1.2.1	Caratterizzazione dei principali elementi di alcune famiglie di dispersione esponenziali univariate notevoli. . . . .	9
1.2.2	Funzione di legame canonica per alcuni modelli lineari generalizzati. . . . .	11
1.2.3	Funzioni di legame canoniche e funzioni di legame disponibili per alcune famiglie di distribuzione indicate con il simbolo $\bullet$ . . . . .	12
2.2.1	Famiglie di legami standardizzati per GLM. . . . .	29
3.3.1	Output per i coefficienti del modello di regressione logistica per il dataset <i>Downs.bc</i> . . . . .	46
3.3.2	Output per i coefficienti di un modello di regressione logistica. . . . .	48
3.3.3	Output per i coefficienti del modello di regressione logistica per il dataset <i>Beetles</i> . . . . .	49
3.3.4	Output del modello di regressione logistica con coda destra modificata per il dataset <i>Beetles</i> con $\psi_1 = 1.92$ e $\eta_0 = 0$ . . . . .	51
3.3.5	Output del modello di regressione logistica con coda sinistra modificata per il dataset <i>Beetles</i> con $\psi_2 = 0.16$ e $\eta_0 = 0$ . . . . .	51





# Introduzione

I modelli lineari generalizzati (GLM) permettono l'adattamento di molti modelli statistici di regressione (Agresti, 2015).

Di primaria importanza è dare una specificazione corretta del modello di regressione partendo dalle sue componenti.

Il legame canonico, il quale lega linearmente la componente aleatoria a quella sistematica all'interno di un modello, è spesso utilizzato poiché ammette una statistica sufficiente minimale per i parametri di regressione e consente una semplice interpretazione del risultato.

In molte applicazioni è stato tuttavia dimostrato che l'utilizzo di modelli di regressione lineari generalizzati con trasformazioni di funzioni di legame parametriche, ha portato, rispetto all'utilizzo di un legame fissato, ad un sostanziale miglioramento nell'adattamento ai dati (Czado, 1997).

Per la definizione di tali funzioni si ricorre a famiglie generatrici che permettono la costruzione di link per molte distribuzioni. Vi sono classi di modelli indicizzate da uno o più parametri, i quali permettono di controllare il comportamento di una o di entrambe le code della funzione di legame determinandone la simmetria.

Il modello di regressione logistica, ad esempio, è largamente utilizzato per modellare la dipendenza di dati binari da variabili esplicative, e questo grazie alla sua semplice forma e interpretazione. Tale modello riesce ad adattarsi in molte applicazioni, tuttavia assume che la curva di probabilità attesa sia simmetrica. Questa restrizione può risultare limitante nel caso si debbano trattare due code asimmetriche per tale curva.

Per risolvere questi problemi si specifica una classe di modelli indicizzata da uno o due parametri in modo tale da comprendere un più ampio spettro di situazioni rispetto al modello logistico.

In questi casi l'interpretazione dei parametri è più complessa se paragonata ad un GLM con il link canonico.

Spesso non si ha sufficiente informazione per l'individuazione corretta della funzione di legame e questo può portare ad errate specificazioni che si traducono, ad esempio, in un aumento di distorsione.

Si ricorre allora all'utilizzo di vari test d'ipotesi al fine di assicurarsi la miglior definizione della funzione di link per il problema in questione.

Fondamentale è riuscire a trovare un bilanciamento tra l'aumento delle va-

---

rianze nelle stime dei coefficienti di regressione ed il miglioramento nell'adattamento del modello ai dati attraverso l'introduzione dei parametri di link. La ricerca della stima di massima verosimiglianza per i parametri in questione è guidata dalla minimizzazione della devianza residua del modello. Verranno infine presentati alcuni esempi di modifiche alle funzioni di legame con conseguente miglioramento nell'adattamento del modello ad dati.

# Capitolo 1

## Modelli Lineari Generalizzati

Dato un insieme di variabili, in molte applicazioni si è interessati allo studio di una variabile dipendente che si cercherà di descrivere attraverso il comportamento di una o più variabili indipendenti. La prima rappresenta la variabile risposta, mentre le seconde prendono il nome di variabili concomitanti, con riferimento a tutte le variabili disponibili nell'insieme di dati, o esplicative, in riferimento ad un dato insieme scelto di concomitanti che entrano a far parte della formula di un modello.

I modelli di regressione vengono utilizzati appunto con tale scopo: spiegare se e come il comportamento di un gruppo di variabili influisce su una variabile di interesse. Il modello di regressione lineare normale (Grigoletto *et al.*, 2017) rappresenta la base per lo sviluppo di modelli più complessi, i modelli lineari generalizzati (McCullagh e Nelder, 1989). Questi ultimi permettono di modellare la relazione tra variabile risposta e covariate quando vengono meno le principali assunzioni su cui si basa il modello lineare normale, rispondendo a particolari problemi distributivi, di non omogeneità della varianza e di non linearità tra media delle risposta e covariate.

### 1.1 Modello di regressione lineare normale

I modelli lineari normali (si veda ad esempio Grigoletto *et al.*, 2017) si compongono di una parte strutturale, la quale spiega la relazione tra la variabile di interesse (o dipendente) e le variabili che ne determinano il comportamento (variabili indipendenti), e di una componente casuale, il termine d'errore, per la quale viene assunta una distribuzione normale. Il modello lineare normale per le osservazioni  $\mathbf{y} = (y_1, \dots, y_n)$  su una risposta quantitativa con variabili esplicative  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ , per  $i = 1 \dots n$ , assume che  $y_i$  sia realizzazione della variabile  $Y_i$ , la quale si lega alle covariate per mezzo della seguente relazione lineare

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \dots, n.$$

In particolare il modello è caratterizzato da tre assunzioni:

1. **Componente aleatoria:** La variabile casuale  $Y_i$  è indipendente dalle altre e si distribuisce come  $Y_i \sim N(\mu_i, \sigma^2)$ , con  $\sigma^2 > 0$  ignoto,  $i = 1, \dots, n$ .  
Con tale assunzione si specifica anche la distribuzione degli errori casuali  $\varepsilon_1, \dots, \varepsilon_n$ , anch'essi indipendenti con media nulla e varianza costante  $\sigma^2$ , con  $\sigma^2 > 0$ :  $\varepsilon_i \sim N(0, \sigma^2)$ ;
2. **Omoschedasticità:**  $Var(\varepsilon_i) = Var(Y_i) = \sigma^2 \quad \forall i$ ;
3. **Componente deterministica:** Si assume che la risposta casuale sia legata alle covariate dalla seguente relazione lineare  $\mu_i = E(Y_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$ ,  $i = 1, \dots, n$ , con  $\beta$  coefficienti di regressione.

Per  $n$  osservazioni, indicando con  $x_{ij}$  il valore assunto per la  $j$ -esima variabile esplicativa per l' $i$ -esimo soggetto, con  $i = 1, \dots, n$ , e  $j = 1, \dots, p$ , la matrice di disegno  $X$ , non stocastica e a rango pieno, è data da

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}.$$

Allora

$$\mathbf{Y} \sim N_n(X\beta, \sigma^2 \mathbb{I}_n), \quad (1.1)$$

con  $E(Y_i) = \mu_i = \eta_i = \mathbf{x}_i \beta$ ,  $Var(Y_i) = \sigma^2 > 0$  e  $\mathbb{I}_n$  matrice identità avente  $n$  righe. Dunque la media  $\mu_i$  è legata al predittore lineare  $\eta_i$  e, fissati i vettori  $\mathbf{x}_i$ , la distribuzione di  $Y = (Y_1, \dots, Y_n)$  è interamente individuata dal parametro  $(p+1)$ -dimensionale  $(\beta_1, \dots, \beta_p, \sigma^2)$ . Nelle applicazioni del modello (1.1), tipicamente la prima colonna di  $X$  è  $(1, \dots, 1)^T$ , ovvero  $\beta_1$  è l'intercetta.

Per ogni fissato  $r \in \{2, \dots, p\}$ , il modello assume che  $E(Y_i)$  aumenti di  $\beta_r$  unità se  $x_{ir}$  viene incrementata di un'unità tenendo inalterati i livelli delle altre esplicative: un coefficiente di regressione esprime dunque una differenza media della risposta tra sottopopolazioni che hanno identici valori di tutte le variabili esplicative, ma differiscono per una unità nella  $r$ -esima. Allora, al netto delle rimanenti  $x_{is}$ , con  $s \neq r$ ,  $\beta_r$  viene anche chiamato *effetto* della corrispondente esplicativa sulla risposta.

### 1.1.1 Funzione di verosimiglianza e stima dei parametri del modello

Il parametro complessivo è  $\theta = (\beta_1, \dots, \beta_p, \sigma^2) \in \Theta = \mathbb{R}^p \times ]0, +\infty[$ . Dalla (1.1) si ottiene la funzione di verosimiglianza

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_p x_{ip})^2\right\} \\ &\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_p x_{ip})^2\right\}. \end{aligned} \quad (1.2)$$

Utilizzando la rappresentazione matriciale del modello (1.1), posto  $\mu = X\beta = \beta_1 \mathbf{x}_1 + \dots + \beta_p \mathbf{x}_p$ , la log-verosimiglianza per  $\theta = (\beta, \sigma^2)$  è

$$\ell(\beta, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta). \quad (1.3)$$

Una stima puntuale per i parametri di regressione  $(\beta_1, \dots, \beta_p)$  è ottenibile con il metodo dei minimi quadrati. Tale stima corrisponde a quella di massima verosimiglianza: per  $\sigma^2$  fissato,  $\ell(\beta, \sigma^2)$  è massima dove è minima la somma dei quadrati degli scarti

$$S(\beta) = (\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta), \quad (1.4)$$

la quale non dipende da  $\sigma^2$ .

Derivando  $S(\beta)$  rispetto a  $\beta$ , ponendo a zero l'equazione ed infine risolvendo sempre rispetto a  $\beta$ , si ottiene la stima di massima verosimiglianza

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}. \quad (1.5)$$

Per la sua stessa natura, lo stimatore di massima verosimiglianza (1.5) è una trasformazione lineare di una variabile aleatoria normale multidimensionale (1.1), e dunque presenta la seguente distribuzione

$$\hat{\beta} \sim N_p(\beta, \sigma^2 (X^T X)^{-1}). \quad (1.6)$$

Sostituendo  $\hat{\beta}$  nell'equazione (1.3), derivando e risolvendo il problema di massimo rispetto a  $\sigma^2$  si ha

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n} = \frac{\mathbf{e}^T \mathbf{e}}{n}, \quad (1.7)$$

dove gli  $e_i = y_i - \hat{y}_i$  sono i residui del modello, con  $\mathbf{e}$  il vettore corrispondente. La distribuzione di tale stimatore è

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-p}^2. \quad (1.8)$$

### 1.1.2 Inferenza sul modello

Con i risultati ottenuti sulla distribuzione degli stimatori dei parametri  $\beta$  si possono definire quantità pivotali per poter fare inferenza su uno o più coefficienti del modello lineare normale. Per esempio, se si è interessati a valutare l'ipotesi di nullità di un gruppo di  $p-p_0$  coefficienti, si può impostare un confronto tra modelli.

Dato il sistema d'ipotesi

$$\begin{cases} H_0 : \beta_{p_0+1} = \beta_{p_0+2} = \dots = \beta_p = 0 \\ H_1 : \exists r \in \{p_0+1, \dots, p\} \text{ tale che } \beta_r \neq 0, \end{cases} \quad (1.9)$$

si stima il modello completo, o corrente con  $p$  regressori e se ne ottengono i residui  $\mathbf{e} = \mathbf{y} - X\hat{\beta}$ . Successivamente si calcola la somma dei quadrati dei residui. Si stima poi il modello ridotto e i residui associati  $\mathbf{e}_0 = \mathbf{y} - X_0\tilde{\beta}_0$ , con  $\tilde{\beta}_0 = (X_0^T X_0)^{-1} X_0^T \mathbf{y}$  e  $X_0$  formata dalle prime  $p_0$  colonne di  $X$ , e se ne calcola la relativa somma dei quadrati.

I modelli sotto ipotesi sono allora annidati. Il sistema d'ipotesi può essere semplificato nel seguente modo

$$\begin{cases} H_0 : \beta_A = 0 \\ H_1 : \beta_A \neq 0, \end{cases} \quad (1.10)$$

dove  $\beta_A$  è il vettore dei parametri  $(\beta_{p_0+1}, \dots, \beta_p)^T$  (mentre si indicherà con  $\beta_B$  il vettore di parametri  $(\beta_1, \dots, \beta_{p_0})^T$ ). Si può allora calcolare il log-rapporto di verosimiglianza

$$\begin{aligned} W_p = W_p(\beta_A)|_{\beta_A=0} &= 2 \left( \ell(\hat{\beta}_B, \hat{\beta}_A, \hat{\sigma}^2) - \ell_0(\tilde{\beta}_0, \tilde{\sigma}^2) \right) = \\ &= 2 \left( -\frac{n}{2} \log \hat{\sigma}^2 + \frac{n}{2} \log \tilde{\sigma}^2 \right) = n \log \frac{\mathbf{e}_0^T \mathbf{e}_0}{\mathbf{e}^T \mathbf{e}}. \end{aligned} \quad (1.11)$$

Si rifiuta l'ipotesi nulla se il valore osservato  $W_p$  è elevato, ovvero se  $\mathbf{e}_0^T \mathbf{e}_0$  è grande rispetto  $\mathbf{e}^T \mathbf{e}$ . Sfruttando l'assunzione di normalità per gli errori, si può ricavare la distribuzione esatta di  $W_p$ : la statistica  $\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} - 1 = \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2}$  è equivalente a  $W_p$  e rappresenta l'aumento relativo della varianza stimata dell'errore che è dovuta al passaggio dal modello completo a quello ridotto. Si ha

$$\frac{n(\tilde{\sigma}^2 - \hat{\sigma}^2)}{\hat{\sigma}^2} \sim \chi_{p-p_0}^2.$$

Sapendo che  $n(\tilde{\sigma}^2 - \hat{\sigma}^2)$  e  $n\hat{\sigma}^2$  sono indipendenti, dalla (1.7) segue

$$F = \frac{\frac{n(\tilde{\sigma}^2 - \hat{\sigma}^2)}{\hat{\sigma}^2(p-p_0)}}{\frac{n\hat{\sigma}^2}{\sigma^2(n-p)}} \sim F_{p-p_0, n-p}. \quad (1.12)$$

Se il valore osservato di  $F$ ,  $F^{oss}$ , è tale che  $F^{oss} > F_{p-p_0, n-p; 1-\alpha}$ , per un dato livello di significatività  $\alpha$ , allora si rifiuta l'ipotesi nulla.

### 1.1.3 Coefficiente di determinazione $R^2$

Al fine di verificare la bontà d'adattamento di un modello di regressione lineare ai dati si fa riferimento al *coefficiente di determinazione*  $R^2$ .

Considerando il modello lineare normale (1.1) si ha  $\mu = X\beta \in V$ , sottospazio vettoriale  $p$ -dimensionale di  $\mathbb{R}^n$  (dove  $p < n$ ). Sia inoltre  $\mathbb{R}^n$  dotato del prodotto scalare  $x^T y = \sum_{i=1}^n x_i y_i$  e della norma  $\|\cdot\|$ , con  $\|y\|^2 = y^T y$ . Se nel modello adottato è presente l'intercetta, e dunque la media campionaria dei valori predetti  $n^{-1}1_n^T \hat{y}$  è uguale alla media campionaria dei valori osservati  $\bar{y} = n^{-1}1_n^T y$ , o, equivalentemente, la media campionaria dei residui è pari a zero, si ha la seguente scomposizione ortogonale per  $y$

$$y = \hat{y} + e,$$

dove  $\hat{y} \in V$  e  $e = \hat{y} - y \in V^\perp$ , con  $V^\perp$  sottospazio vettoriale ortogonale a  $V$  in quanto l'elemento  $\hat{y}$  è la proiezione ortogonale di  $y$  su  $V$ . Allora vale

$$\|y - \bar{y}1_n\|^2 = \|\hat{y} - \bar{y}1_n\|^2 + \|y - \hat{y}\|^2,$$

o, equivalentemente

$$SQT = SQ_{reg} + SQ_{res},$$

dove  $SQT$  è la devianza totale, somma di due quantità non negative: la devianza spiegata dal modello di regressione,  $SQ_{reg}$ , e la devianza residua, ovvero la somma dei quadrati dei residui del modello,  $SQ_{res}$ .

Allora l'indice  $R^2$  è dato dal rapporto

$$R^2 = \frac{SQ_{reg}}{SQT} = 1 - \frac{SQ_{res}}{SQT}. \quad (1.13)$$

Esso è dunque il rapporto tra la devianza spiegata dal modello e la devianza totale. Per sua costruzione, esso è compreso tra 0 e 1; quando è prossimo allo zero si ha che la distribuzione della variabile di interesse non dipende, o dipende debolmente, dalle variabili concomitanti. Al contrario se  $R^2$  è prossimo a 1 il modello si adatta bene ai dati e fornisce valori predetti prossimi a quelli osservati.

### 1.1.4 Residui e bontà d'adattamento

Infine, per condurre una verifica empirica delle assunzioni del modello, si esegue un'analisi dei residui.

Si considerano alcune analisi grafiche al fine di verificare le ipotesi di omoschedasticità, normalità ed incorrelazione tra residui e covariate. In particolare si studiano i seguenti diagrammi sul piano cartesiano:

- *Diagramma dei residui:*  
( $i, e_i$ ),  $i = 1, \dots, n$ . Tale grafico permette di evidenziare la presenza di correlazione tra le osservazioni. Affinché il modello possa essere considerato valido, esso non deve evidenziare alcun andamento sistematico;

- *Diagramma dei residui rispetto le esplicative del modello:*  
 $(x_{ij}, e_i) i = 1, \dots, n, j = 1, \dots, p$ . Tale grafico permette di evidenziare eventuali correlazioni tra residui ed esplicative che, in base alle assunzioni del modello, devono essere assenti;
- *Diagramma dei residui rispetto ai valori stimati dal modello:*  
 $(\hat{y}_i, e_i) i = 1, \dots, n$ . Tale grafico, come il precedente, al fine di garantire un buon adattamento del modello ai dati non deve evidenziare alcun andamento sistematico.

Per testare l'ipotesi di normalità, ci si può basare sul confronto tra le funzioni di ripartizione teorica ed empirica dei residui, accostando a tale risultato il *test di Kolmogorov-Smirnov*, oppure si possono confrontare i quantili teorici ed empirici mediante il *grafico quantile-quantile* (Bortot *et al.*, 2000) affiancando a questo il *test di Shapiro-Wilk* (Ventura e Racugno, 2017).

## 1.2 Modello di regressione lineare generalizzato

Possono esserci situazioni in cui le assunzioni del modello lineare normale vengono meno. Si passa allora alla specificazione di una più ampia famiglia di modelli che consente di porre rimedio agli scostamenti delle ipotesi di normalità, omoschedasticità e di linearità tra media della risposta ed esplicative. A tal fine potrebbero esser impiegate trasformazioni della variabile risposta, come la lognormale, la trasformazione di Box-cox o la trasformazione della radice quadrata (Grigoletto *et al.*, 2017), per il trattamento della non linearità, non normalità ed eteroschedasticità rispettivamente. Tuttavia, la trasformazione della variabile risposta potrebbe rivelarsi inadeguata nel momento in cui si desidera avere un modello per  $Y$ , e non per una sua trasformazione, oppure, quando la risposta è discreta e non vi è trasformazione che conduca alla normalità.

I modelli lineari generalizzati (si veda ad esempio Agresti, 2015) per risposte univariate rappresentano un'estensione del modello di regressione lineare normale.

In particolare si assume che ogni componente del vettore della variabile risposta  $Y$  abbia distribuzione appartenente alla *famiglia esponenziale*, e che una funzione del valore atteso della risposta sia funzione lineare delle variabili esplicative.

### 1.2.1 Famiglie di dispersione esponenziale

Nei GLM (*Generalized Linear Models*) si assume che la funzione di densità (o di probabilità) della risposta appartenga ad un modello parametrico a sua volta incluso in una più ampia classe, la classe delle *famiglie di dispersione esponenziale* (Nelder e Wedderburn, 1972).



La densità di  $Y_i$ ,  $i = 1, \dots, n$ , appartiene a una *famiglia di dispersione esponenziale univariata* se è esprimibile nella forma

$$p(y_i; \theta_i, \phi) = \exp \left\{ \frac{\theta_i y_i - b(\theta_i)}{a_i(\phi)} + c(y_i; \phi) \right\}, \quad (1.14)$$

con  $y_i \in \mathcal{Y}$ ,  $\theta_i \in \Theta$ ,  $b(\cdot)$  e  $c(\cdot)$  funzioni note,  $a_i(\phi) > 0$ , con  $\mathcal{Y}$  supporto della variabile risposta e  $\Theta$  spazio parametrico. Il parametro  $\theta_i$  è detto *parametro naturale*, mentre  $\phi$  è detto *parametro di dispersione*. Talvolta si ha  $a_i(\phi) = \phi/\omega_i$  con  $\phi > 0$  e  $\omega_i$ ,  $i = 1, \dots, n$ , pesi noti.

Specificando nella precedente formula l'espressione delle funzioni  $a_i(\cdot)$ ,  $b(\cdot)$  e  $c(\cdot)$ , si ottiene un particolare modello parametrico con una specifica funzione della media  $v(\mu_i)$  e spazio delle medie  $\mathbb{M}$ ; si trovano per esempio le distribuzioni normale, gamma, binomiale e Poisson (Tabella 1.2.1).

Tabella 1.2.1: Caratterizzazione dei principali elementi di alcune famiglie di dispersione esponenziali univariate notevoli.

Modello	$N(\mu_i, \sigma^2)$	$Ga(\alpha, \alpha/\mu_i)$	$\frac{1}{m_i}Bi(m_i, \mu_i)$	$P(\mu_i)$
$\mathcal{Y}$	$\mathbb{R}$	$[0, +\infty)$	$\{1/m_i, \dots, 1\}$	$\mathbb{N}$
$\theta_i$	$\mu_i$	$-1/\mu_i$	$\log \left( \frac{\mu_i}{1-\mu_i} \right)$	$\log(\mu_i)$
$\Theta$	$\mathbb{R}$	$(-\infty, 0)$	$\mathbb{R}$	$\mathbb{R}$
$b(\theta_i)$	$\theta_i^2/2$	$-\log(-\theta_i)$	$\log(1 + \exp(\theta_i))$	$\exp(\theta_i)$
$\phi$	$\sigma^2$	$1/\alpha$	1	1
$a_i(\phi)$	$\phi$	$\phi$	$1/m_i$	1
$\mathbb{M}$	$\mathbb{R}$	$(0, +\infty)$	$(0, 1)$	$(0, +\infty)$
$v(\mu_i)$	1	$\mu_i^2$	$\mu_i(1 - \mu_i)$	$\mu_i$

### 1.2.2 Media e Varianza

Per tutti i modelli parametrici della forma (1.14) si possono ottenere risultati generali, come le espressioni di media e varianza.

Sia  $Y_i$  con densità (1.14) e sia  $\ell(\theta_i, \phi; y_i)$  la funzione di log-verosimiglianza per  $(\theta_i, \phi)$  basata su  $y_i$ ,  $i = 1, \dots, n$ . Risulta

$$\ell(\theta_i, \phi; y_i) = \frac{\theta_i y_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi),$$

con derivata prima rispetto  $\theta_i$  data da

$$\frac{\partial \ell(\theta_i, \phi)}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a_i(\phi)}$$

e derivata parziale seconda

$$\frac{\partial^2 \ell(\theta_i, \phi)}{\partial \theta_i^2} = \frac{-b''(\theta_i)}{a_i(\phi)},$$

dove si sono indicate con  $b'(\theta_i)$  e  $b''(\theta_i)$  le prime due derivate di  $b(\theta_i)$ .  
In modelli con verosimiglianza regolare, dove  $\mathcal{Y}$  non dipende da  $\theta$ , vale

- i.  $E\left(\frac{\partial \ell(\theta_i, \phi; Y_i)}{\partial \theta_i}\right) = 0;$
- ii.  $E\left(\left(\frac{\partial \ell(\theta_i, \phi; Y_i)}{\partial \theta_i}\right)^2\right) = \frac{Var(Y_i)}{(a_i(\phi)^2)} = \frac{b''(\theta_i)}{a_i(\phi)}.$

Si ottiene così

- $E(Y_i) = b'(\theta_i)$ , indipendente da  $\phi$ ;
- $Var(Y_i) = a_i(\phi)b''(\theta_i) = a_i(\phi)v(\mu_i)$ , con  $v(\mu_i) = b''(\theta_i)|_{\theta_i=\theta_i(\mu_i)}$  funzione di varianza.

Poiché  $Var(Y_i) > 0$  per ogni  $\theta_i$ , la funzione  $\mu_i(\cdot)$  è monotona crescente con dominio  $\Theta$  e codominio lo spazio delle medie  $\mathbb{M}$ . Conoscendo  $\mathbb{M}$  e  $v(\mu_i)$  è possibile risalire a  $b(\theta_i)$ , la quale, nota la funzione  $a_i(\phi)$ , definisce uno specifico modello parametrico nella classe dei modelli con densità (1.14). Allora la funzione di varianza assieme al suo dominio caratterizza uno specifico modello nella famiglia di dispersione esponenziale:

$$Y_i \sim DE_1(\mu_i, a_i(\phi)v(\mu_i)), \quad \mu_i \in \mathbb{M}, i = 1, \dots, n. \quad (1.15)$$

### 1.2.3 Componenti di un modello lineare generalizzato

Un modello lineare generalizzato viene descritto dalle seguenti tre componenti:

1. **La componente aleatoria:** per ogni osservazione la variabile casuale  $Y_i$  è indipendente dalle altre e si distribuisce nel seguente modo:  $Y_i \sim DE_1(\mu_i, a_i(\phi)v(\mu_i));$
2. **La componente sistematica:** per ogni osservazione si definisce il *predittore lineare*  $\eta_i$  nel seguente modo:  $\eta_i = \mathbf{x}_i\beta$  con  $\beta \in \mathbb{R}^p$  vettore dei parametri di regressione;
3. **La funzione di legame:** è la funzione  $g(\cdot)$  monotona crescente e dunque invertibile, che collega la media della componente casuale a quella sistematica:  
 $g(E(Y_i)) = g(\mu_i) = \eta_i = \mathbf{x}_i\beta$ . Tale funzione garantisce la compatibilità dei domini di  $E(Y_i)$  e  $\eta_i$ ,  $g : \mathbb{M} \rightarrow \mathbb{R}$ .

Poiché la varianza di  $Y_i$  risulta

$$\text{Var}(Y_i) = a_i(\phi)v(\mu_i) = a_i(\phi)v(g^{-1}(\mathbf{x}_i\beta)), \quad (1.16)$$

nell'ambito dei modelli lineari generalizzati sono ammesse particolari forme di eteroschedasticità.

### 1.2.4 Funzione di legame

La funzione di legame è particolarmente conveniente quando lo spazio delle medie  $\mathbb{M}$  non coincide con  $\mathbb{R}$ , come accade per le distribuzioni binomiale, Poisson e gamma. In tal modo è possibile modellare in termini lineari la dipendenza della variabile risposta, il cui supporto è un sottoinsieme proprio di  $\mathbb{R}$ , dalle variabili esplicative.

Per ciascuna specificazione del modello statistico di dispersione esponenziale per una data risposta, fra tutte le possibili funzioni di legame  $g(\cdot)$  è privilegiata la funzione  $g(\cdot) = g(\mu_i) = g(E(Y_i))$  tale che

$$\begin{aligned} \theta &= \theta(\mu_i) = \eta_i = \mathbf{x}_i\beta \\ \Rightarrow g(E(Y_i)) &= g(\mu_i) = \theta_i = \theta_i(\mu_i), \end{aligned} \quad (1.17)$$

con la quale il parametro naturale  $\theta_i$  della  $DE_1$  risulta combinazione lineare delle variabili esplicative con coefficienti  $\beta$ ,  $\theta_i = \mathbf{x}_i\beta$ ,  $i = 1, \dots, n$ .

La (1.17) è detta *funzione di legame canonica* (Tabella 1.2.2).

Tabella 1.2.2: Funzione di legame canonica per alcuni modelli lineari generalizzati.

Modello	$N(\mu_i, \sigma^2)$	$Ga(\alpha, \alpha/\mu_i)$	$\frac{1}{m_i}Bi(m_i, \mu_i)$	$P(\mu_i)$
legame canonico	$\mu_i$	$-1/\mu_i$	$\log\left(\frac{\mu_i}{1-\mu_i}\right)$	$\log(\mu_i)$

Ogni famiglia di distribuzione può essere formulata facendo riferimento a più funzioni di legame diverse. La scelta viene fatta in base allo scopo dello studio e ai dati disponibili. (Tabella 1.2.3).

La scelta della funzione di legame è importante per poter avere un modello più flessibile. Un GLM con funzione di legame canonica presenta una funzione di verosimiglianza concava, assicura l'esistenza di una statistica sufficiente minimale per i parametri di regressione<sup>1</sup> ed una semplice interpretazione dei

<sup>1</sup>Sia  $\theta$  parametro sul quale si vuole fare inferenza e  $y$  vettore di osservazioni disponibili. Una statistica è una trasformazione dei dati  $t = f(y)$ , con  $f$  non iniettiva. Si dice che  $s$  è statistica sufficiente per l'inferenza su  $\theta$  se  $p_Y(y; \theta) = p_S(s; \theta)p_{Y|S=s}(y)$ , dove  $p_{Y|S=s}(y)$  non dipende da  $\theta$ . In altre parole,  $s$  è sufficiente per l'inferenza su  $\theta$  se e solo se la distribuzione di  $Y|S = s$  non è inferente su  $\theta$ . Tale statistica permette dunque una riduzione dei dati e del modello senza perdita di informazione su  $\theta$  (Pace e Salvani, 2001).

Tabella 1.2.3: Funzioni di legame canoniche e funzioni di legame disponibili per alcune famiglie di distribuzione indicate con il simbolo •

$g(\mu)$	legame	Distribuzione			
		<i>binomial</i>	<i>Gamma</i>	<i>gaussian</i>	<i>poisson</i>
$\log(\frac{\mu}{1-\mu})$	<i>logit</i>	<i>canonica</i>			
$\Phi^{-1}(\mu)$	<i>probit</i>	•			
$\log(-\log(1-\mu))$	<i>cloglog</i>	•			
$\mu$	<i>identity</i>		•	<i>canonica</i>	•
$\mu^{-1}$	<i>inverse</i>		<i>canonica</i>	•	
$\log(\mu)$	<i>log</i>	•	•	•	<i>canonica</i>
$\mu^{1/2}$	<i>sqrt</i>				•

parametri di regressione: in tal caso si ottiene un modello lineare per il parametro naturale della famiglia di distribuzione esponenziale. Per esempio la funzione di legame logistica permette una semplice interpretazione del predittore lineare in termini del logaritmo della quota, il quale facilita a sua volta l'interpretazione del risultato.

### 1.2.5 Verosimiglianza e inferenza

Siano  $Y_1, \dots, Y_n$  variabili casuali distribuite secondo le seguenti assunzioni:

- $Y_1, \dots, Y_n$  v.c. univariate indipendenti;
- $g(E(Y_i)) = g(\mu_i) = \eta_i = \mathbf{x}_i\beta$ ;
- $Y_i \sim DE_1(\mu_i, a_i(\phi)v(\mu_i))$ ,  $\mu_i \in \mathbb{M}$ .

Allora  $\mathbf{Y} = (Y_1, \dots, Y_n)$  ha densità congiunta data dal prodotto delle densità marginali (1.14) e la funzione di log-verosimiglianza per  $(\beta, \phi)$  risulta

$$\ell(\beta, \phi) = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + \sum_{i=1}^n c(y_i, \phi), \quad (1.18)$$

con  $\theta_i = \theta(\mu_i) = \theta(g^{-1}(\mathbf{x}_i\beta))$ .

Per il calcolo delle derivate della log-verosimiglianza sono necessarie le derivate del parametro naturale rispetto ai parametri di regressione. In generale, se  $g(\mu_i) = \mathbf{x}_i\beta = \eta_i$ , allora  $\mu_i = g^{-1}(\mathbf{x}_i\beta) = g^{-1}(\eta_i)$  e dunque  $\theta_i = \theta_i(\mu_i) = \theta_i(g^{-1}(\mathbf{x}_i\beta))$ . Per la regola di derivazione della funzione inversa

si ha

$$\frac{\partial \theta_i}{\partial \mu_i} = \theta'_i(\mu_i) = \frac{1}{\mu'_i(\theta_i)}|_{\theta_i=\theta_i(\mu_i)} = \frac{1}{b''(\theta_i)}|_{\theta_i=\theta_i(\mu_i)}.$$

Inoltre sapendo che  $v(\mu_i) = b''(\theta_i)|_{\theta_i=\theta_i(\mu_i)}$  si ha

$$\theta'_i(\mu_i) = \frac{1}{v(\mu_i)}.$$

Dato  $\mu_i = g^{-1}(\eta_i)$ , si ha

$$\frac{\partial \theta_i}{\partial \eta_i} = \frac{1}{g'(\mu_i)}|_{\mu_i=g'(\eta_i)}.$$

Con la regola di derivazione delle funzioni composte si ottiene allora

$$\frac{\partial \theta_i}{\partial \beta_r} = \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_r} = \frac{1}{v(\mu_i)} \frac{\partial \mu_i}{\partial \beta_r}, \quad r = 1, \dots, p, \quad (1.19)$$

dove  $\mu_i = g^{-1}(\eta_i)$ . Inoltre, se  $g(\cdot)$  è la funzione di legame canonica,  $g(\mu_i) = \theta_i(\mu_i)$ , risulta  $g'(\mu_i) = 1/v(\mu_i)$ , da cui si ha

$$\frac{\partial \theta_i}{\partial \beta_r} = x_{ir}, \quad r = 1, \dots, p. \quad (1.20)$$

Il vettore delle derivate della log-verosimiglianza rispetto ai parametri (vettore *score* o punteggio) ha componenti

$$\ell_r = \frac{\partial \ell(\beta, \phi)}{\partial \beta_r} = \sum_{i=1}^n \frac{1}{a_i(\phi)} \left( y_i \frac{\partial \theta_i}{\partial \beta_r} - \frac{\partial b(\theta_i)}{\partial \beta_r} \right), \quad r = 1, \dots, p, \quad (1.21)$$

$$\ell_\phi = \frac{\partial \ell(\beta, \phi)}{\partial \phi} = - \sum_{i=1}^n \frac{a'_i(\phi)}{(a_i(\phi))^2} (y_i \theta_i - b(\theta_i)) + \sum_{i=1}^n c'(y_i, \phi), \quad (1.22)$$

dove  $a'_i(\phi) = \partial a_i(\phi)/\partial \phi$  e  $c'(y_i, \phi) = \partial c(y_i, \phi)/\partial \phi$ . Allora

$$\frac{\partial b(\theta_i)}{\partial \beta_r} = b'(\theta_i) \frac{\partial \theta_i}{\partial \beta_r} = \mu_i \frac{\partial \theta_i}{\partial \beta_r},$$

e quindi le componenti relative a  $\beta$  del vettore *score* possono essere poste nella forma

$$\ell_r = \sum_{i=1}^n \frac{(y_i - \mu_i)}{a_i(\phi)} \frac{\partial \theta_i}{\partial \beta_r}, \quad r = 1, \dots, p. \quad (1.23)$$

Pertanto, dalla (1.19) e assumendo  $\phi$  noto, si hanno le *equazioni di verosimiglianza* per  $\beta$

$$\ell_r = \sum_{i=1}^n \frac{(y_i - \mu_i)}{Var(Y_i)} \frac{\partial \mu_i}{\partial \beta_r} = 0, \quad r = 1, \dots, p. \quad (1.24)$$

Se la funzione di legame è quella canonica, allora  $g'(\mu_i) = 1/v(\mu_i)$  e le equazioni di verosimiglianza (1.23) si semplificano nel seguente modo

$$\sum_{i=1}^n \frac{y_i x_{ir}}{a_i(\phi)} = \sum_{i=1}^n \frac{\mu_i x_{ir}}{a_i(\phi)}, \quad r = 1, \dots, p. \quad (1.25)$$

Le equazioni precedenti possono essere riscritte in forma matriciale. In particolare la (1.24) è

$$(y - \mu)^T V^{-1} D = 0, \quad (1.26)$$

con  $(y - \mu)^T = (y_1 - \mu_1, \dots, y_n - \mu_n)$ ,  $V = \text{diag}(\text{Var}(Y_i))$ ,  $i = 1, \dots, n$  e  $D$  matrice  $n \times p$ ,  $D = \text{diag}(g'(\mu_i)^{-1})X$ . Dal momento che, al di fuori del modello lineare normale per cui si ha  $\mu = X\beta$ ,  $D = X$  e  $V^{-1} = (\sigma^2)^{-1}I_n$ , le equazioni (1.26) non hanno soluzione esplicita; queste andranno risolte attraverso metodi iterativi.

### 1.2.6 Informazione osservata e attesa

Derivando le (1.23) rispetto a  $\phi$  si ha

$$\ell_{r\phi} = \frac{\partial^2 \ell(\beta, \phi)}{\partial \beta_r \partial \phi} = - \sum_{i=1}^n \frac{a'_i(\phi)}{(a_i(\phi))^2} (y_i - \mu_i) \frac{\partial \theta_i}{\partial \beta_r}, \quad r = 1, \dots, p,$$

e si nota che hanno valore atteso nullo essendo  $E(Y_i) = \mu_i$ . In primo luogo si nota che  $\beta$  e  $\phi$  sono *parametri ortogonali*, ovvero il blocco  $i_{\beta\phi}$  dell'informazione attesa ha tutti elementi uguali a zero. Come conseguenza dell'ortogonalità si ha che gli stimatori di massima verosimiglianza di  $\beta$  e  $\phi$  sono asintoticamente indipendenti e allora, per l'inferenza su  $\beta$ , è sufficiente il blocco dell'informazione osservata e attesa relativa a  $\beta$ . Derivando allora rispetto a  $\beta_s$  la (1.23) si ha

$$j_{rs} = -\ell_{rs} = \sum_{i=1}^n \frac{1}{a_i(\phi)} \left[ \frac{\partial \mu_i}{\partial \beta_s} \frac{\partial \theta_i}{\partial \beta_r} - (y_i - \mu_i) \frac{\partial^2 \theta_i}{\partial \beta_r \partial \beta_s} \right]. \quad (1.27)$$

Se il legame è canonico  $\partial^2 \theta_i / (\partial \beta_r \partial \beta_s) = 0$  e dunque l'espressione di  $j_{rs}$  risulta

$$j_{rs} = \sum_{i=1}^n \frac{1}{a_i(\phi)} \frac{\partial \mu_i}{\partial \beta_s} \frac{\partial \theta_i}{\partial \beta_r},$$

che è una quantità non stocastica e coincide con il proprio valore atteso. Se il legame non è canonico si ottiene la seguente espressione per l'informazione attesa

$$i_{rs} = \sum_{i=1}^n \frac{1}{a_i(\phi)} \frac{x_{ir} x_{is}}{(g'(\mu_i))^2 v(\mu_i)} \quad (1.28)$$

ed in forma matriciale si ha

$$i_{\beta\beta} = X^T W X, \quad (1.29)$$

dove

$$W = \text{diag}(w_i), \quad \text{con} \quad w_i = \frac{1}{(g'(\mu_i))^2 \text{Var}(Y_i)}, \quad i = 1, \dots, n. \quad (1.30)$$

Se il legame è canonico si ha

$$w_i = \frac{1}{(1/v(\mu_i))^2 a_i(\phi) v(\mu_i)} = \frac{v(\mu_i)}{a_i(\phi)}.$$

Il risultato generale per la normalità asintotica dello stimatore di massima verosimiglianza fornisce l'approssimazione

$$\hat{\beta} \sim N_p(\beta, (X^T \hat{W} X)^{-1}), \quad (1.31)$$

per  $n$  elevato. Allora  $(X^T \hat{W} X)^{-1}$  è una stima consistente per la matrice di varianza e covarianza di  $\beta$ , dove  $\hat{W}$  indica la matrice  $W$  calcolata con  $\beta = \hat{\beta}$  e per  $\phi$ , se ignoto, pari ad una sua stima consistente non distorta ottenibile attraverso il metodo dei momenti come

$$\tilde{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{v(\mu_i)}. \quad (1.32)$$

### 1.2.7 Minimi quadrati pesati iterati

Le equazioni di verosimiglianza (1.24) non ammettono in genere soluzione esplicita; vanno allora risolte mediante dei metodi iterativi come il metodo di Newton-Raphson. Posto  $\ell_*$  il vettore con elementi  $\ell_r$  e  $j = j_{\beta\beta}$  la matrice di informazione osservata con elementi  $-\ell_{rs}$ , la  $(m+1)$ -esima iterazione fornisce l'approssimazione

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + \left( j(\hat{\beta}^{(m)}) \right)^{-1} \ell_*(\hat{\beta}^{(m)}).$$

La matrice di informazione osservata può essere sostituita con il suo valore atteso  $i_{\beta\beta}$ , in tal modo si mantiene la convergenza dell'algoritmo e le espressioni risultano semplificate

$$i_{\beta\beta}(\hat{\beta}^{(m)}) \hat{\beta}^{(m+1)} = i_{\beta\beta}(\hat{\beta}^{(m)}) \hat{\beta}^{(m)} + \ell_*(\hat{\beta}^{(m)}). \quad (1.33)$$

Sfruttando la relazione (1.29), la (1.33) diventa

$$X^T W X \hat{\beta}^{(m+1)} = X^T W z^{(m)}, \quad (1.34)$$

con le quantità  $W$  e  $z^{(m)}$  valutate in  $\hat{\beta}^{(m)}$ . La

$$z^{(m)} = X\hat{\beta}^{(m)} + u,$$

è la *variabile dipendente aggiustata* ed una sua generica componente è

$$z_i^{(m)} = \mathbf{x}_i\hat{\beta}^{(m)} + (y_i - \mu_i)g'(\mu_i), \quad i = 1, \dots, n.$$

La (1.34) coincide formalmente con l'espressione delle equazioni normali per gli stimatori dei minimi quadrati generalizzati. La  $(m+1)$ -esima iterazione dell'algoritmo calcola  $\hat{\beta}^{(m+1)}$  come stima dei minimi quadrati generalizzati in un modello lineare avente come matrice del modello  $X$ , come variabile risposta  $z^{(m)}$  (calcolata in  $\hat{\beta}^{(m)}$ ) e come matrice dei pesi  $W$  (anch'essa calcolata in  $\hat{\beta}^{(m)}$ ). Poiché la matrice dei pesi varia da iterazione a iterazione, tale algoritmo iterativo è detto dei *minimi quadrati pesati iterati*.

### 1.2.8 Devianza e bontà d'adattamento

Il processo d'adattamento di un modello ai dati può esser visto come un modo di sostituire un insieme di dati osservati  $y$  con uno di dati stimati  $\hat{\mu}$  da un modello con un numero ridotto di parametri. Dal momento che in genere le stime non coincidono perfettamente con i valori osservati la questione si concentra su tale discrepanza.

Misure di discrepanza o di bontà d'adattamento di modelli differenti possono esser costruite a partire dal logaritmo del rapporto delle funzioni di verosimiglianza, detto *devianza residua*. Date  $n$  osservazioni possiamo adattare un modello contenente fino a  $n$  parametri. Il modello più semplice, il *modello nullo*, è quello ottenibile sotto l'ipotesi di omogeneità ed ha un solo parametro (l'intercetta); tale modello richiede allora che le  $\mu_i$  siano tutte costanti e quindi indipendenti dalle covariate, e ripone tutta la variabilità delle  $y_i$  alla componente casuale. All'estremo opposto si ha il *modello saturo* con  $n$  parametri, ciascuno per ogni osservazione, il quale porta ad avere un adattamento perfetto ad dati e ad avere tutte le medie che variano senza alcun legame tra loro. Tale modello pone tutta la variabilità nella componente sistematica.

Il modello nullo è dunque troppo semplice, mentre quello saturo non ci dà alcuna informazione in quanto non riassume i dati. Tuttavia quest'ultimo ci fornisce termine di confronto per misurare la discrepanza con un modello intermedio contenente  $p$  parametri.

Siano  $Y_1, \dots, Y_n$  variabili casuali indipendenti aventi distribuzione marginale  $DE_1(\mu_i, a_i(\phi)v(\mu_i))$ , con  $g(\mu_i) = \mathbf{x}_i\beta$  e  $a_i(\phi) = \phi/\omega_i$ ,  $i = 1, \dots, n$ . Scelto un modello, sia  $\hat{\theta}_i$  la stima di massima verosimiglianza per il parametro naturale  $\theta_i$  corrispondente alla stima della media  $\hat{\mu}_i$ . Sia  $\hat{\theta}_i$  la stima di



$\theta_i$  per il modello saturo con corrispondente  $\tilde{\mu}_i = y_i$ , per  $i = 1, \dots, n$ .

Per studiare la capacità esplicativa del modello corrente si considera il test log-rapporto di verosimiglianza che pone a confronto tale modello con quello saturo.

Data la funzione di verosimiglianza per  $(\theta, \phi)$

$$\ell(\theta, \phi) = \sum_{i=1}^n \left\{ \frac{\theta_i y_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi) \right\}, \quad (1.35)$$

si definisce devianza residua la seguente quantità

$$D(y; \hat{\mu}) = 2\phi(\ell(y; \tilde{\mu}, \phi) - \ell(y; \hat{\mu}, \phi)) = 2\phi(\ell(y; y, \phi) - \ell(y; \hat{\mu}, \phi)). \quad (1.36)$$

La quantità

$$D(y; \hat{\mu})/\phi \quad (1.37)$$

è detta devianza residua riscalata del modello corrente.

Si utilizza il seguente test log-rapporto di verosimiglianza

$$W = D(y; \hat{\mu})/\phi. \quad (1.38)$$

Come regola empirica, se  $W \leq n - p$ , o poco più grande, allora il modello corrente si adatta bene ai dati osservati.

Si possono inoltre apportare semplificazioni al modello corrente. Indicando con  $D(y; \hat{\mu}_0)$  la devianza residua del modello ridotto e con  $D(y; \hat{\mu})$  quella del modello corrente, per verificare la possibilità di ridurre quest'ultimo ad un sottomodulo in esso annidato, con  $p_0 < p$  variabili esplicative, si utilizza la statistica test log-rapporto di verosimiglianza

$$W_P = \frac{D(y; \hat{\mu}_0) - D(y; \hat{\mu})}{\phi}, \quad (1.39)$$

con distribuzione asintotica nulla  $\chi_{p-p_0}^2$ . Valori grandi di  $W_P$  portano a rifiutare l'ipotesi di semplificazione.

Se  $\phi$  è ignoto, la devianza riscalata stimata  $D(y; \hat{\mu})/\tilde{\phi}$  è approssimativamente uguale a  $n - p$  e dunque non può essere utilizzata come test per la bontà d'adattamento; si ricorre alla statistica  $\{D(y; \hat{\mu}_0) - D(y; \hat{\mu})\}/\{\tilde{\phi}(p - p_0)\}$  con  $\tilde{\phi}$  data da (1.32). Tale statistica ha distribuzione approssimata  $F$  con  $(p - p_0, n - p)$  gradi di libertà (Bortot *et al.*, 2000).

Infine, anche il test per il modello nullo contro il modello corrente è un caso particolare della statistica (1.39), e precisamente è

$$W_P = \frac{D(y; \hat{\mu}_{null}) - D(y; \hat{\mu})}{\phi},$$

con distribuzione asintotica  $\chi_{p-1}^2$ , dove  $D(y; \hat{\mu}_{null})$  è la *devianza nulla*, o devianza del modello minimale.

### 1.2.9 Residui

Una volta adattato il modello lineare generalizzato si passa a confrontare i valori stimati  $\hat{\mu}_i$  con le osservazioni  $y_i$ , per  $i = 1, \dots, n$ . Tale confronto è possibile mediante un'ispezione grafica dei residui, i quali sono strumenti validi per identificare osservazioni anomale o influenti, per valutare la scelta della funzione di varianza, della funzione di legame o dei termini inclusi nel predittore lineare, e dunque in generale, per verificare eventuali allontanamenti dalle ipotesi del modello. A tale scopo è importante che i residui abbiano un comportamento simile a quello dei residui standardizzati nel modello lineare normale: in particolare, se il modello è stato correttamente specificato, questi devono avere media nulla, varianza unitaria e distribuzione prossima alla normale standard.

Tuttavia nei GLM la componente d'errore non è esplicitamente definita; inoltre le varianze dei residui possono essere molto diverse tra loro e questo porta ad avere più definizioni. Vi sono due tipi di residui che vengono usati per gli scopi sopra citati: i *residui di Pearson* ed i *residui di devianza*.

L'estensione diretta del concetto di residuo standardizzato è data dal residuo di Pearson. Assumendo  $a_i(\phi) = \phi/\omega_i$ , esso è definito come

$$r_i^P = \frac{y_i - \hat{\mu}_i}{v(\hat{\mu}_i)/\omega_i}, \quad i = 1, \dots, n. \quad (1.40)$$

Tali residui sono così chiamati in quanto, nel modello di Poisson in cui  $v(\mu_i) = \mu_i$  e  $\omega_i = 1$ , la statistica

$$X^2 = \sum_{i=1}^n (r_i^P)^2 = \sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i},$$

è nota come statistica di Pearson, ed equivale alla devianza nel modello con legame canonico; infatti essa è un indice della distanza tra modello stimato e i dati, ed i suoi addendi misurano quanto questa distanza è attribuibile a ciascuna osservazione.

I residui di devianza si definiscono invece come la radice con segno del contributo della singola osservazione a  $D(y; \hat{\mu})$ , ossia

$$r_i^D = \text{sgn}(y_i - \hat{\mu}_i) \sqrt{(D_i)}, \quad i = 1, \dots, n, \quad (1.41)$$

dove

$$D_i = 2\omega_i \{y_i[\theta_i(y_i) - \theta_i(\hat{\mu}_i)] - [b(\theta_i(y_i)) - b(\theta_i(\hat{\mu}_i))]\}, \quad i = 1, \dots, n.$$

Esistono inoltre le forme standardizzate per i due tipi di residui sopra definiti. I *residui di Pearson standardizzati* sono

$$r_i^{PS} = \frac{r_i^P}{\sqrt{\tilde{\phi}(1 - \hat{h}_{ii}^\omega)}} = \frac{y_i - \hat{\mu}_i}{\sqrt{\tilde{\phi}v(\hat{\mu}_i)(1 - \hat{h}_{ii}^\omega)/\omega_i}}, \quad i = 1, \dots, n,$$

dove  $Var(Y_i - \hat{\mu}_i) \doteq Var(Y_i)(1 - h_{ii}^\omega)$ , con  $h_{ii}^\omega$  elemento  $i$ -esimo sulla diagonale della matrice  $H$  generalizzata (Agresti, 2015), data da

$$H_\omega = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2},$$

con  $W$  data dalla (1.30).

Analogamente si definiscono i *residui di devianza standardizzati* come

$$r_i^{DS} = \frac{r_i^D}{\sqrt{\tilde{\phi}(1 - \hat{h}_{ii}^\omega)}}, \quad i = 1, \dots, n.$$

Per verificare la correttezza del modello specificato si studiano i grafici dei residui rispetto i valori stimati  $\hat{\mu}_i$  e rispetto le variabili esplicative; da tale osservazione, per avere un modello soddisfacente, non devono emergere andamenti sistematici.

### 1.3 Conclusioni

Nel prossimo capitolo verrà approfondito il ruolo della funzione di legame all'interno dei GLM: i vantaggi che una specificazione corretta porta all'adattamento del modello, i limiti della funzione di legame canonica e particolari estensioni che consentono di ridurre la distorsione delle stime dei parametri migliorando l'adattamento ai dati.



## Capitolo 2

# Funzioni di legame parametriche

Nei GLM la funzione di legame canonica è quella che ammette una maggiore semplicità nell'interpretazione del risultato, ed assicura l'esistenza di una statistica sufficiente minimale per i parametri di regressione.

Ad esempio, nel caso di un modello con risposta binaria, la funzione di legame canonica è la logistica. Il modello di regressione logistica permette l'interpretazione del predittore lineare in termini del logaritmo della quota (dove per quota s'intende il rapporto tra la probabilità di successo di un evento ed il suo complemento a uno). Il parametro  $r$ -esimo di regressione,  $\beta_r$ , all'interno del predittore lineare, esprime quindi l'effetto sul logaritmo della quota di un incremento unitario della corrispondente covariata  $x_{ir}$ , con le altre  $x_{is}$ , con  $r \neq s$ , costanti. Allora nel caso di un'esplicativa continua, tale incremento si traduce in una moltiplicazione della quota corrispondente di un fattore dato da  $\exp(\beta_r)$ , mentre nel caso di un'esplicativa dicotomica l'incremento si osserva nel logaritmo del rapporto delle quote, il quale indicherà che il rapporto tra la probabilità di successo ed insuccesso per i dati trattati ( $x_{ir} = 1$ ), sarà  $\exp(\beta_r)$  volte il medesimo rapporto per i dati non trattati ( $x_{ir} = 0$ ).

Possono però esserci situazioni dove la funzione di legame può essere erroneamente specificata. Questo fa sì che essa non fornisca il miglior adattamento ad un insieme di dati; l'errata specificazione è plausibile nel momento in cui la precisazione della funzione avviene prima di avere informazioni sufficienti riguardo la scelta della stessa, e ciò porta ad avere una sostanziale distorsione nelle stime dei parametri di regressione e della media della risposta (Flach, 2014).

Per far fronte a tale problema si può far dipendere la funzione di legame da uno o più parametri, scelti sulla base della bontà d'adattamento. Esistono

diverse famiglie di legami parametrici, come la famiglia ad un parametro di Prentice (Prentice, 1976) e le famiglie a due parametri di Pregibon (Pregibon, 1980), Stukel e Czado (Czado, 1994). L'utilizzo di tali modelli apporta un miglioramento dal punto di vista della bontà d'adattamento ma complica la questione dal lato dell'interpretazione dei parametri di regressione.

Va inoltre precisato che, se da un lato, a confronto con un GLM con legame canonico, l'utilizzo di famiglie di legami parametrici può migliorare l'adattamento in quanto può dar luogo ad un modello più flessibile, dall'altro si nota che la distorsione nell'adattamento ai dati viene ridotta a spese della varianza. Si ha infatti un aumento delle varianze delle stime dei coefficienti di regressione, e tale costo, che cresce nel caso in cui questi parametri sono correlati con quello di legame, può esser ridotto a zero attraverso una loro parametrizzazione ortogonale (Czado e Santner, 1992). L'obiettivo è allora quello di trovare un bilanciamento tra questi due aspetti.

## 2.1 GLM con link parametrici

Le funzioni di legame parametriche vengono scelte:

1. in modo da includere la funzione di legame canonica del modello considerato;
2. sulla base del confronto di adattamenti con la massima verosimiglianza negli insiemi di dati;
3. per la loro flessibilità;
4. per la loro semplicità dal punto di vista matematico.

Sulla base del primo punto si considera la funzione di legame canonica entro una più ampia classe di legami  $\mathcal{F} = \{F(\cdot, \psi) : \psi \in \Psi\}$  indicizzata da un parametro  $\psi$ ; il legame canonico si ha nel caso in cui  $\psi = \psi_0$ .

Allora la definizione estesa dei modelli lineari generalizzati con link parametrici riprende quella proposta nella Sezione 1.2.3, generalizzando però la componente di legame al punto 3. nel seguente modo:

3. Il predittore lineare  $\eta_i(\beta)$  è collegato alla media  $\mu_i$  di  $Y_i$  da  $\mu_i = F(\eta_i(\beta), \psi)$ , per una qualche  $F(\cdot, \psi)$  in  $\mathcal{F} = \{F(\cdot, \psi) : \psi \in \Psi\}$ .

Per un fissato parametro  $\psi$  si ritorna alla classe dei modelli lineari generalizzati, mentre questo non è più vero se tale parametro è ignoto e deve esser stimato congiuntamente a quelli di regressione.

Siano ora le componenti di un GLM con link parametrico definite nel seguente modo:

- componente casuale:  $Y_i$  con densità  

$$p(y_i; \theta_i, \phi) = \exp \left\{ \frac{\theta_i y_i - b(\theta_i)}{a_i(\phi)} + c(y_i; \phi) \right\}, \quad i = 1, \dots, n;$$
- componente sistematica:  $\eta_i(\beta) = \mathbf{x}_i \beta$ , con  $\beta$  parametri di regressione, per  $i = 1, \dots, n$ ;
- componente di legame parametrica:  $\mu_i = F(\eta_i(\beta), \psi)$ , per una qualche  $F(\cdot, \psi)$  in  $\mathcal{F} = \{F(\cdot, \psi) : \psi \in \Psi\}$ , con  $i = 1, \dots, n$ .

Nel seguito si farà riferimento all'informazione di Fisher per la valutazione dell'ortogonalizzazione dei parametri nei GLM con link parametrici. L'informazione di Fisher  $I(\delta)$  per il modello appena definito, con  $\delta = (\beta, \phi)$ , è espressa nel seguente modo

$$I(\delta) = a(\phi)^{-1} \begin{pmatrix} I_{\beta, \beta} & I'_{\beta, \psi} \\ I_{\beta, \psi} & I'_{\psi, \psi} \end{pmatrix}, \quad (2.1)$$

dove  $I_{\beta, \beta}$  è una matrice di dimensioni  $(p+1) \times (p+1)$  con elementi

$$(I_{\beta, \beta})_{rs} = \sum_{i=1}^n x_{is} x_{ir} F_{i1}^2 d_i, \quad r \geq 0, \quad s \geq 0,$$

$I_{\beta, \psi}$  è un vettore  $(p+1)$  con

$$(I_{\beta, \psi})_r = \sum_{i=1}^n x_{ir} F_{i1} F_{i2} d_i, \quad r \geq 0,$$

e  $I_{\psi, \psi}$  è un numero reale dato da

$$I_{\psi, \psi} = \sum_{i=1}^n F_{i2}^2 d_i.$$

Dove  $F_{i1} = \frac{\partial}{\partial \eta} F(\eta, \psi)|_{\eta=\eta_i}$  e  $F_{i2} = \frac{\partial}{\partial \psi} F(\eta, \psi)|_{\eta=\eta_i}$  per  $i = 1, \dots, n$  e  $d_i = \frac{d\theta_i}{d\mu_i}$ .

### 2.1.1 Ortogonalizzazione dei parametri nei GLM con link parametrici

Siano  $\psi$  scalare, e  $\beta$  vettore. Una parametrizzazione  $(\psi, \beta)$  è detta *localmente ortogonale* in  $(\psi^0, \beta^0)$  (Cox e Reid, 2013), se l'informazione di Fisher attesa in  $(\psi^0, \beta^0)$  è zero per la combinazione di parametri  $(\psi, \beta)$ . La parametrizzazione è detta *globalmente ortogonale* se è localmente ortogonale per ogni valore del parametro  $(\psi, \beta)$ .

L'ortogonalizzazione globale porta a due conseguenze:

1. le stime di massima verosimiglianza  $\hat{\psi}$  e  $\hat{\beta}$  per  $(\psi, \beta)$  sono asintoticamente indipendenti;
2. lo standard error per  $\hat{\beta}$  (o per  $\hat{\psi}$ ) è lo stesso che si ha con  $\psi$  (o  $\beta$ ) noto o ignoto.

La prima proprietà implica che l'ortogonalità tra il parametro di legame  $\psi$  e i parametri di regressione  $\beta$  nei GLM con link parametrici riduce asintoticamente a zero l'aumento della varianza di  $\hat{\beta}$ . Un altro vantaggio derivante dall'utilizzo della parametrizzazione ortogonale sta nel fatto che essa induce una riduzione del tempo computazionale nel calcolo dei parametri  $\beta$  e  $\psi$ , e di conseguenza porta anche ad una maggiore stabilità numerica.

In generale,  $(\psi, \beta)$  non rappresenta una parametrizzazione ortogonale, dunque bisogna cercarne una  $(\psi, \lambda)$ . Una condizione sufficiente per l'ortogonalità di  $\psi$  e  $\beta$  (Cox e Reid, 2013) è

$$\frac{\partial}{\partial \psi} F(\eta_i(\beta), \psi) = 0, \quad \text{per } i = 1, \dots, n. \quad (2.2)$$

In generale tale condizione non vale per ogni  $\psi$  e  $\beta$  dal momento che questo equivale a forzare  $F(\cdot, \psi)$  ad essere indipendente da  $\psi$ . L'obiettivo è dunque quello di trovare una famiglia di legame parametrica  $F(\cdot, \psi)$  in modo tale che il parametro di legame  $\psi$  sia ortogonale al parametro  $\beta$ . Tale restrizione è ragionevole dal momento che una riparametrizzazione da  $(\psi, \beta)$  a  $(\psi, \lambda(\psi, \beta))$  dipenderebbe anche dalla matrice di disegno  $X$ , e dunque, i nuovi parametri  $\lambda(\psi, \beta)$  perderebbero, altrimenti, la loro semplice interpretazione di parametri di regressione.

Se la condizione (2.2) porta ad avere  $\eta_0$ , allora si ha una ortogonalizzazione locale in  $(\psi, \beta^0)$  con predittore lineare  $\eta_i(\beta^0) = \eta_0$ . Inoltre, se si ha che i parametri di regressione sono nulli (ad eccezione dell'intercetta  $\beta_0$ ), e dunque  $\beta_1 = \beta_2 = \dots = \beta_p = 0$ , allora  $(\psi, \beta^0 = (\eta_0, \beta_1 = \dots = \beta_p = 0))$  è tale per cui  $\eta(\beta^0) = \eta_0$ . Sotto condizioni di regolarità  $\frac{\partial}{\partial \psi} F(\eta_0, \psi) = 0$  equivale ad avere  $\mu_0$  tale per cui  $F(\eta_0, \psi) = \mu_0 \forall \psi \in \Psi$ .

Avendo chiaro i vantaggi derivanti dall'uso di una parametrizzazione ortogonale, si è interessati ad avere una ortogonalizzazione locale almeno in un intorno di  $\eta_0$ . A tal fine si considera l'approssimazione del primo ordine di Taylor per  $F(\cdot, \psi)$  in  $\eta_0$ :

$$F(\eta, \psi) \simeq F(\eta_0, \psi) + \frac{\partial}{\partial \eta} F(\eta, \psi)|_{\eta=\eta_0} (\eta - \eta_0). \quad (2.3)$$

Se  $F(\eta, \psi)|_{\eta=\eta_0}$  è indipendente da  $\psi$ , allora la (2.3) mostra che  $F(\eta, \psi)$  è approssimativamente localmente indipendente da  $\psi$  in un intorno di  $\eta_0$ . Questo implica l'ortogonalità approssimata di  $(\psi, \beta)$  se  $\beta$  è tale che  $\eta(\beta)$  è contenuto in un intorno di  $\eta_0$  nel quale vale l'approssimazione (2.3).

L'ortogonalizzazione approssimata attorno a  $\eta_0$  è preferibile rispetto all'ortogonalizzazione locale in  $\eta_0$  dal momento che i contributi alla matrice di



informazione attesa, e quindi all'aumento della varianza, di punti  $\eta(\beta)$  che cadono entro un intorno di  $\eta_0$  sono trascurabili. Mentre questo non è garantito dall'ortogonalizzazione locale in  $\eta_0$ . Infine, i valori di  $\beta$  per i quali  $\eta(\beta)$  è contenuto in tale intorno, dipendono dal particolare valore di  $\psi$  e dalla particolare matrice di disegno  $X$ .

Allora si studiano le famiglie di legame  $\mathcal{F} = \{F(\cdot, \psi), \psi \in \Psi\}$  che soddisfano le seguenti condizioni:

- i.  $\exists \eta_0$  tale che  $F(\eta_0, \psi) = \mu_0 \quad \forall \psi$ ;
- ii.  $\exists \frac{\partial}{\partial \eta} F(\eta, \psi)|_{\eta=\eta_0} = s_0 \quad \forall \psi$ .

### 2.1.2 Standardizzazione di famiglie di link parametrici nei GLM

Un altro aspetto da considerare nella scelta di una particolare famiglia di legame è il seguente: la famiglia di legame  $\mathcal{F} = \{F(\cdot, \psi), \psi \in \Psi\}$  deve poter essere invariante rispetto trasformazioni di posizione e di scala.

Allora una famiglia  $\mathcal{F}$  verrà detta invariante per scala e posizione se le trasformazioni di scala e posizione per  $\eta$  sono tali da assicurare che esiste un'unica  $F \in \mathcal{F}$  che può venir scelta per una data posizione e scala delle covariate. Questo significa che

$$\begin{aligned} \text{per ogni } F(\cdot, \psi) \in \mathcal{F} \text{ la funzione } G_{a,b}(\eta, \psi) = F(a + b\eta, \psi) \\ \text{è contenuta in } \mathcal{F} \text{ per } a = 0 \text{ e } b = 1 \text{ o } 0. \end{aligned} \quad (2.4)$$

Tale condizione viene anche vista come un vincolo d'identificabilità. Dal momento che  $G_{a,b}(\frac{\eta_0 - a}{b}, \psi) = F(\eta_0, \psi) = \mu_0$ , la (2.4) è soddisfatta per  $\eta^* = \frac{\eta_0 - a}{b}$  per ogni  $a$  e  $b \neq 0$ . Inoltre, per  $b \neq 1$ , si ha

$$\frac{\partial}{\partial \eta} G_{a,b}(\eta, \psi)|_{\eta=\eta^*} = \frac{\partial}{\partial \eta} F(a + b\eta, \psi)|_{\eta=\eta^*} = b \frac{\partial}{\partial \eta} F(\eta, \psi)|_{\eta=\eta_0} = bs_0 \neq s_0.$$

Si può dunque passare ad una definizione riassuntiva; la famiglia di legame  $\mathcal{F} = \{F(\cdot, \psi), \psi \in \Psi\}$  è detta:

1.  $\eta_0$ -invariante di posizione se  $\exists$  un valore  $\eta_0$  tale che  $F(\eta_0, \psi) = s_0 \quad \forall \psi \in \Psi$ ;
2.  $(\eta_1 - \eta_2)$ -invariante di scala se  $\exists$  due valori  $\eta_1$  e  $\eta_2$  tali che  $F(\eta_1, \psi) = s_1$  e  $F(\eta_2, \psi) = s_2$  sono indipendenti da  $\psi$ ;
3.  $\eta_0$ -standardizzato se  $\exists$  un valore  $\eta_0$  tale che  $F(\eta_0, \psi) = s_0$  e  $\frac{\partial}{\partial \eta} F(\eta, \psi)|_{\eta=\eta_0}$  è indipendente da  $\psi \quad \forall \psi \in \Psi$ .

Il primo punto riprende la (i), il secondo la (2.4) mentre il terzo la (ii).

Nella definizione precedente non vengono scelti particolari valori per  $\eta_0$ ,  $\eta_1$  o  $\eta_2$ ; tuttavia è ragionevole assumere  $\eta_0 \leq \eta_1 \leq \eta_2$ . Dal momento che

la scelta delle posizioni  $(\eta_1, \eta_2)$  influenza le possibili forme delle funzioni di legame contenute nella famiglia  $\mathcal{F}$ , è difficile fare una scelta ragionevole per loro in anticipo.

Per una data famiglia di legami, la scelta di  $\eta_0$  determina quella di  $\mu_0$  e  $s_0$ . L'assunzione  $s_0 = F(\eta_0, \psi)$  per ogni  $\psi$ , equivale all'ortogonalizzazione locale di  $(\psi, \beta_0 = \eta_0, \beta = 0)$ . Inoltre, conoscendo i vantaggi derivanti dall'ortogonalizzazione, è di interesse combinare gli effetti derivanti da quest'ultima e dalla standardizzazione: usando delle covariate centrate  $\eta_i(\beta) = \beta_0 + \beta_1(x_{i1} - \bar{x}_1) + \dots + (x_{ip} - \bar{x}_p)$ , si ha che queste variano attorno a  $\beta_0$ . Allora si sceglie  $\eta_0 = \beta_0$ . Se  $\mu_0$  è la media di  $Y$  quando si hanno covariate nulle, questo porta ad avere  $s_0 = F(\eta_0, \psi) = F(\beta_0, \psi) = \mu_0$ . Dal momento che in generale  $\mu_0$  è ignoto, si sceglie  $s_0 = \hat{\mu}_0$  dove  $\hat{\mu}_0$  è una stima di  $\mu$ .

## 2.2 Famiglie parametriche a uno e due parametri

Si supponga ora di avere una famiglia di legami  $\mathcal{G} = \{G(\cdot, \psi), \psi \in \Psi\}$  che è  $(\mu_0 = 0, s_0 = 1)$ -standardizzata per  $\eta_0 = 0$ .

Esistono famiglie generatrici che permettono la costruzione di link per molte distribuzioni e in particolare per quella binomiale. Le principali sono le seguenti (Czado, 1997):

$$G(\eta, \psi) = (1 + \psi\eta)^{1/\psi}, \quad (2.5)$$

$$G(\eta, \psi) = \frac{\exp(\eta\psi) - 1}{\psi}, \quad (2.6)$$

$$G(\eta, \psi) = \frac{\log(1 + \psi\eta)}{\psi}, \quad (2.7)$$

$$G(\eta, \psi) = \frac{(\eta + 1)^\psi - 1}{\psi}, \quad (2.8)$$

$$G(\eta, \psi) = (1 + \psi \exp(\eta))^{1/\psi}. \quad (2.9)$$

Quest'ultima classe tuttavia non è  $(\mu_0, s_0)$ -standardizzata per ogni valore finito di  $\eta_0$ , ma solo per  $\eta_0 \rightarrow \infty$ .

Più in generale vi sono due approcci per costruire una famiglia di legame  $(\mu_0, s_0)$ -standardizzata in  $\eta_0$ .

- i. Si considera la famiglia  $F(\eta, \psi) = G(s_0(\eta - \eta_0), \psi) + \mu_0$ , dove  $G$  è assunta essere  $(\mu_0 = 0, s_0 = 1)$ -standardizzata in  $\eta_0$ . Questo implica che  $G(0, \psi) = 0$  e  $\frac{\partial}{\partial \eta} G(\eta, \psi)|_{\eta=0} = 1$ . Da questa segue

$$F(\eta_0, \psi) = G(0, \psi) + \mu_0 = 0 + \mu_0 = \mu_0 \text{ e}$$

$$\frac{\partial}{\partial \eta} F(\eta, \psi)|_{\eta=\eta_0} = G(s_0(\eta - \eta_0), \psi)|_{\eta=\eta_0} = s_0 \frac{\partial}{\partial \eta} G(\eta, \psi)|_{\eta=0} = s_0.$$

Allora  $\mathcal{F} = \{F(\eta, \psi) = G(s_0(\eta - \eta_0), \psi) + \mu_0, \psi \in \Psi\}$  è una famiglia di legame  $(\mu_0, s_0)$ -standardizzata in  $\eta_0$ .

- ii. In alternativa si possono utilizzare trasformazioni monotone. Si consideri  $M(\cdot)$  con  $M(\eta^*) = \mu_0$  e  $M'(\eta^*) = s_0$ . Si definisce  $F(\eta, \psi) = M(G(\eta - \eta_0, \psi) + \eta^*)$ ,

allora  $F(\eta_0, \psi) = M(G(0, \psi) + \eta^*) = M(0, \eta^*) = M(\eta^*) = \mu_0$ , e

$$\frac{\partial}{\partial \eta} F(\eta, \psi)|_{\eta=\eta_0} = M'(G(0, \psi) + \eta^*) \frac{\partial}{\partial \eta} G(\eta - \eta_0, \psi)|_{\eta=\eta_0} = M'(\eta^*) = s_0.$$

### 2.2.1 Famiglie di trasformazioni di potenza

Si consideri ora la (2.8) la quale permette una modifica di entrambe le code, anche separatamente, del predittore  $\eta$ . Vi sono funzioni di trasformazioni di potenza che possono essere applicate a più distribuzioni all'interno dei GLM, nel seguito saranno considerate le più note.

Sia una famiglia di trasformazioni di potenza  $h(\cdot)$  (Fahrmeir *et al.*, 1992) e sia  $\mathcal{F} = \{F(\cdot, \psi), \psi \in \Psi\}$  una famiglia di link parametrici indicizzata dal parametro  $\psi$ . Allora  $h(\cdot)$  è data da

$$h_{\eta_0}(\eta, \psi) = \begin{cases} \eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1} & \text{se } \eta \geq \eta_0 \text{ e } \psi_1 \neq 0 \\ \eta_0 + \log(\eta - \eta_0 + 1) & \text{se } \eta \geq \eta_0 \text{ e } \psi_1 = 0 \\ \eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2} & \text{se } \eta < \eta_0 \text{ e } \psi_2 \neq 0 \\ \eta_0 - \log(-\eta + \eta_0 + 1) & \text{se } \eta < \eta_0 \text{ e } \psi_2 = 0 \end{cases} \quad (2.10)$$

In particolare, tale specificazione di  $h(\cdot)$  ammette una modifica di entrambe le code (Figura 2.2.1). Una funzione che permette la modifica della sola coda

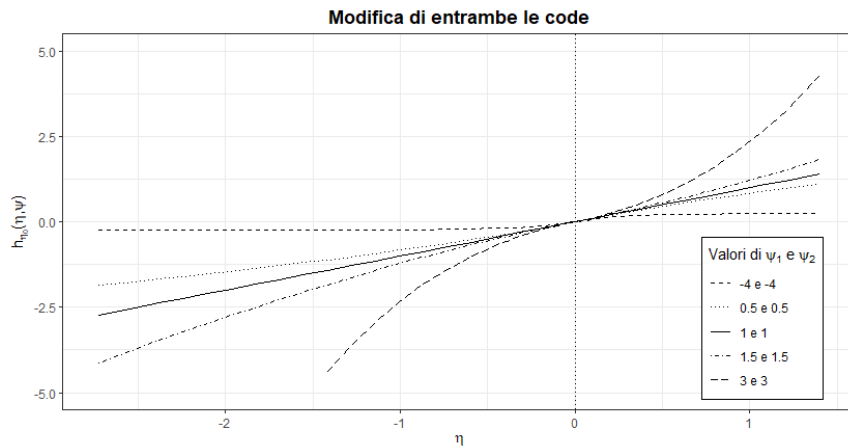


Figura 2.2.1: Modifica di entrambe le code per diversi valori di  $\psi_1$  e  $\psi_2$  con  $\eta_0 = 0$ .

destra è (Figura 2.2.2)

$$h_{\eta_0}(\eta, \psi = \psi_1) = \begin{cases} \eta_0 + \log(\eta - \eta_0 + 1) & \text{se } \eta \geq \eta_0 \text{ e } \psi_1 = 0 \\ \eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1} & \text{se } \eta \geq \eta_0 \text{ e } \psi_1 \neq 0 \\ \eta & \text{se } \eta < \eta_0, \end{cases} \quad (2.11)$$

mentre, per la sola coda sinistra si ha (Figura 2.2.3)

$$h_{\eta_0}(\eta, \psi = \psi_2) = \begin{cases} \eta & \text{se } \eta \geq \eta_0 \\ \eta_0 - \log(-\eta + \eta_0 + 1) & \text{se } \eta < \eta_0 \text{ e } \psi_2 = 0 \\ \eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2} & \text{se } \eta < \eta_0 \text{ e } \psi_2 \neq 0. \end{cases} \quad (2.12)$$

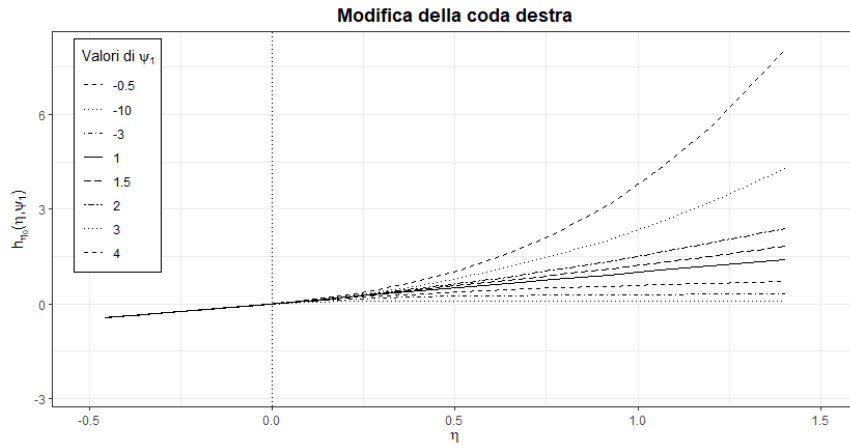


Figura 2.2.2: Modifica della coda destra per diversi valori di  $\psi_1$  con  $\eta_0 = 0$ .

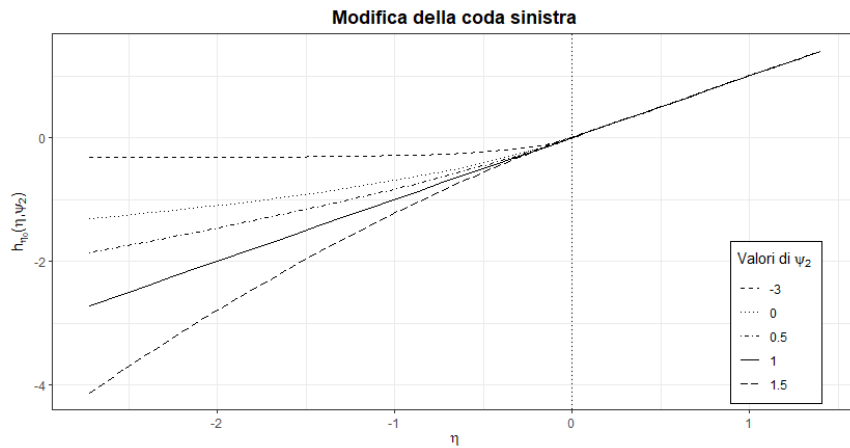


Figura 2.2.3: Modifica della coda sinistra per diversi valori di  $\psi_2$  con  $\eta_0 = 0$ .

Il parametro  $\psi_1$  va a modificare la coda destra (e dunque i valori maggiori di un  $\eta_0$  fissato) del predittore lineare, mentre  $\psi_2$  quella sinistra (i valori minori di  $\eta_0$ ).

Si osserva che  $h_{\eta_0}(\eta, \psi) = \eta_0$  e  $\frac{\partial}{\partial \psi} h_{\eta_0}(\eta, \psi)|_{\eta=\eta_0}$  è indipendente da  $\psi$ . Allora  $h_{\eta_0}(\eta, \psi)$  ed in generale, ogni funzione di  $h_{\eta_0}(\eta, \psi)$ , corrisponde ad un legame  $\eta_0$ -standardizzato. Allora se si pone  $\psi = (1, \psi_2)$  oppure  $\psi = (\psi_1, 1)$  si ottengono rispettivamente le modifiche per la sola coda sinistra e destra. Per  $\psi_i < 1$ , con  $i = 1, 2$ , si ha una diminuzione della pendenza della curva, mentre si ha un aumento per  $\psi_i > 1$ ,  $i = 1, 2$ .

Per giustificare il comportamento di tali funzioni quando i parametri assumono valore nullo si considera il limite per  $\psi_i \rightarrow 0$  di  $\eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_i} - 1}{\psi_i}$ . Applicando il teorema di de l'Hospital (Bertsch *et al.*, 2007) si arriva a quanto formulato nelle equazioni (2.2.1), (2.2.2) e (2.2.3).

Per  $\psi = 1$ ,  $h_{\eta_0}(\eta, \psi)$  è la funzione identità. Ricordando che si vuole che la funzione di legame canonica sia inclusa all'interno della famiglia di legame  $\mathcal{F} = \{F(\cdot, \psi), \psi \in \Psi\}$ , si presentano ora nella Tabella 2.2.1 le famiglie di legame  $\eta_0$ -standardizzate per le distribuzioni più note.

Tabella 2.2.1: Famiglie di legami standardizzati per GLM.

Distribuzione	Legame canonico $F(\eta)$	Famiglia di legame $\mathcal{F} = \{F(\cdot, \psi), \psi \in \Psi\}$	Modifiche ammesse
Normale	$F(\eta) = \eta$	$F(\eta, \psi) = h_{\eta_0}(\eta, \psi)$	Entrambe
Binomiale	$F(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)}$	$F(\eta, \psi) = \frac{\exp h_{\eta_0}(\eta_0, \psi)}{1 + \exp h_{\eta_0}(\eta_0, \psi)}$	Entrambe
Poisson	$F(\eta) = \exp(\eta)$	$F(\eta, \psi) = \exp h_{\eta_0}(\eta, \psi)$	Destra
Gamma	$F(\eta) = \eta^{-1}$	$F(\eta, \psi) = \frac{1}{h_{\eta_0}(\eta, \psi)}$	Sinistra

### 2.2.2 Famiglie di trasformazioni per la regressione binaria

I modelli di regressione binomiali classici, come quello logistico o quello che utilizza il link probit, possono essere estesi al fine di includere famiglie di trasformazioni di legame. Il fine è sempre il medesimo: migliorare l'adattamento ai dati ed evitare errate specificazioni del legame.

Molte delle famiglie generatrici sopra descritte vengono dunque utilizzate per trattare dati provenienti da una distribuzione binomiale: è quindi di interesse studiare la differenza che una trasformazione del link canonico, ovvero quello logistico, apporta all'adattamento del modello.

La (2.5) viene usata per la creazione di una famiglia simmetrica, la famiglia simmetrica di Aranda-Ordaz (Aranda-Ordaz, 1981). Tale famiglia è

così definita

$$g(\theta, \psi) = \frac{2\theta^\psi - (1-\theta)^\psi}{\psi\theta^\psi + (1-\theta)^\psi}, \quad (2.13)$$

dove  $\theta \in (0, 1)$  rappresenta la probabilità di successo e  $\psi$  il parametro di trasformazione. Dal momento che  $G(\theta, \psi) = -G(1-\theta, \psi)$  e  $G(\theta, \psi) = G(\theta, -\psi)$ , e dunque trattando successi ed insuccessi in modo simmetrico, tale famiglia  $G(\theta, \psi)$  è simmetrica.

Si osserva che per  $\psi \rightarrow 0$ , la (2.13) ritorna la trasformazione logistica, mentre per  $\psi = 1$  si ha una semplice trasformazione lineare.

Una famiglia di trasformazioni asimmetrica, invece, è

$$g(\theta, \psi) = \log\left(\frac{(1-\theta)^{-\psi} - 1}{\psi}\right), \quad (2.14)$$

la quale corrisponde alla funzione logit per  $\psi = 1$  e alla log-log complementare per  $\psi \rightarrow 0$ .

La famiglia sviluppata da Pregibon (Pregibon, 1980), definita a partire dalla famiglia generatrice (2.5), è così specificata

$$g(\theta, \psi_1, \psi_2) = \frac{(\theta)^{\psi_1 - \psi_2} - 1}{\psi_1 - \psi_2} - \frac{(1-\theta)^{\psi_1 + \psi_2} - 1}{\psi_1 + \psi_2}. \quad (2.15)$$

Per  $\psi_1 = \psi_2 \rightarrow 0$  si torna ad avere il legame logistico. Il parametro  $\psi_1$  controlla la pesantezza delle code, ed in particolare al diminuire di  $\psi_1$  il peso delle code aumenta; mentre  $\psi_2$  regola l'asimmetria della distribuzione. Per  $\psi_2 = 0$  si ha dunque una distribuzione simmetrica.

Nell'ambito della regressione binaria, un'altra famiglia a un parametro che si ottiene a partire dalla (2.9) è la famiglia parametrica di Burr-Prentice (Burr, 1942). Data la famiglia

$$F(\eta, \psi) = 1 - \frac{1}{(1 + \exp(\eta))^\psi}, \quad \text{per } \eta \in \mathbb{R} \text{ e } \psi > 0, \quad (2.16)$$

si considera l'ortogonalizzazione locale la quale consente di passare dal modello di regressione binaria definito come

$$p(x) = 1 - \frac{1}{(1 + \exp\{\beta_0 + \beta^T x\})^\psi}, \quad (2.17)$$

al modello

$$p(x) = 1 - \frac{1}{(1 + (\lambda_0^{1/\psi} - 1) \exp\{\lambda^T x\})^\psi} \equiv H(\lambda^T x, \psi, \lambda_0), \quad (2.18)$$

trasformando  $\beta_0 = \log(\lambda_0^{1/\psi} - 1)$  per  $\lambda_0 > 1$  e  $\psi > 0$ , e  $\beta_j = \lambda_j$  per  $j \in \{1, \dots, p\}$ . La (2.18) è dunque la famiglia a un parametro di Burr.

Si nota che tale famiglia è sovrapparametrizzata nel momento in cui  $\lambda = 0$ , in quanto tutti gli eventi avranno la stessa probabilità  $H(0, \psi, \lambda_0)$  che dipende da due parametri. Questo dimostra che  $(\psi, \lambda_0)$  non sono identificabili e che la matrice di informazione di Fisher degenera per  $\lambda = 0$ .

Con il passaggio dal modello (2.17) non ortogonalizzato al modello (2.18) non si hanno cambiamenti per quanto riguarda l'interpretazione dei parametri di regressione; la differenza risiede nelle stime puntuali ed intervallari di  $\beta_0$  e  $\lambda_0$ .

Infine, come analizzato per le famiglie precedentemente citate, la famiglia  $H(\eta, \psi, \lambda_0)$  include il legame logistico per  $\psi = 1$ , la trasformazione log-log complementare per  $\psi \rightarrow \infty$ , e una distribuzione a code pesanti per  $\psi \rightarrow 0$ . La sola distribuzione simmetrica che si ritrova è la logistica, mentre si hanno distribuzioni con asimmetria positiva per  $\psi < 1$  e negativa per  $\psi > 1$ .

Vi è infine la famiglia definita da Stukel (Stukel, 1989), il quale ha utilizzato le (2.6) e (2.7) per definire un insieme di trasformazioni per entrambe le code del link parametrico. Allora, sempre per il modello logistico, la formula generale proposta, è la seguente

$$h_\psi(\eta) = \log(\mu/(1 - \mu)), \quad (2.19)$$

dove  $h_\psi(\eta)$  è una funzione di  $\eta$  non lineare strettamente decrescente indicizzata da due parametri di forma,  $\psi_1$  e  $\psi_2$ . Tale funzione è così definita per  $\eta \geq 0$  (o per  $\mu \geq \frac{1}{2}$ )

$$h(\eta, \psi_1) = \begin{cases} \psi_1^{-1}(\exp(\psi_1|\eta|) - 1) & \text{se } \psi_1 > 0 \\ \eta & \text{se } \psi_1 = 0 \\ -\psi_1^{-1} \log(1 - \psi_1|\eta|) & \text{se } \psi_1 < 0, \end{cases} \quad (2.20)$$

mentre per  $\eta \leq 0$  (o per  $\mu \leq \frac{1}{2}$ ) si ha

$$h(\eta, \psi_2) = \begin{cases} -\psi_2^{-1}(\exp(\psi_2|\eta|) - 1) & \text{se } \psi_2 > 0 \\ \eta & \text{se } \psi_2 = 0 \\ \psi_2^{-1} \log(1 - \psi_2|\eta|) & \text{se } \psi_2 < 0. \end{cases} \quad (2.21)$$

La media assume dunque valori differenti a seconda della funzione  $h$  (Figura 2.2.5). Con  $\psi_1 = \psi_2 = 0$  si ha  $h(\eta) = \eta$  e dunque il modello logistico. Per valori diversi da zero di  $\psi_1$  e  $\psi_2$  si hanno funzioni che crescono più velocemente ( $\psi_1, \psi_2 > 0$ ) e più lentamente ( $\psi_1, \psi_2 < 0$ ) rispetto alla curva logistica. Ogni parametro è responsabile del comportamento di una delle code; inoltre se  $\psi_1 = \psi_2$  come in Figura 2.2.4, allora la corrispondente curva  $\mu(\eta)$  è simmetrica, al contrario, le code vengono trattate in modo asimmetrico.

Da questa famiglia a due parametri si può ottenere una classe ad un unico

parametro per modelli simmetrici quando  $\psi = \psi_1 = \psi_2$ . Questa classe può venir utilizzata per valutare eventuali scostamenti (simmetrici) dal modello logistico. Il modello simmetrico è  $\text{logit}(\mu) = \eta + \frac{1}{2}\psi\eta^2 \cdot \text{sgn}(\eta)$ . Quando  $\psi_1$  (oppure  $\psi_2$ ) è nullo, la classe ad un parametro indicizzata da  $\psi_2$  (o  $\psi_1$ ), può essere usata per affinare l'adattamento della coda inferiore (o superiore). La famiglia  $h_{\psi_2}$  ha asimmetria positiva per  $\psi_2 > 0$  e negativa per  $\psi_2 < 0$ ; analogamente la famiglia  $h_{\psi_1}$  presenta asimmetria negativa per  $\psi_1 > 0$  e positiva per  $\psi_1 < 0$ .

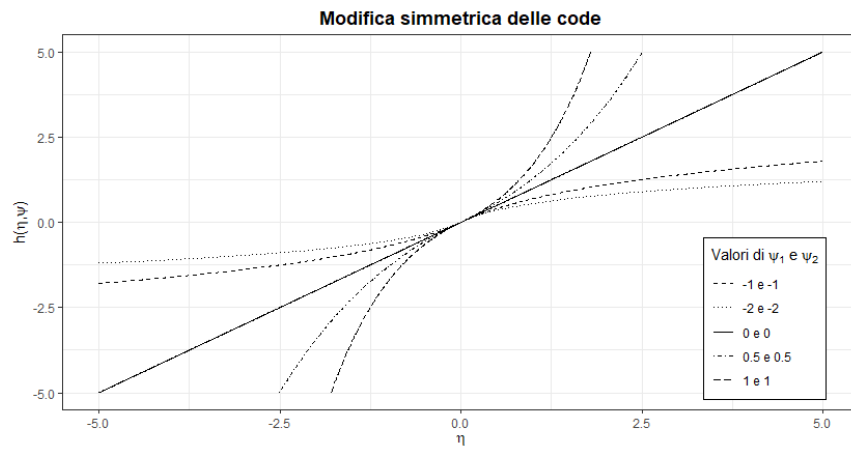


Figura 2.2.4: Modifica simmetrica delle code per diversi valori di  $\psi_1$  e  $\psi_2$  con la famiglia di trasformazioni di Stukel per dati binari.

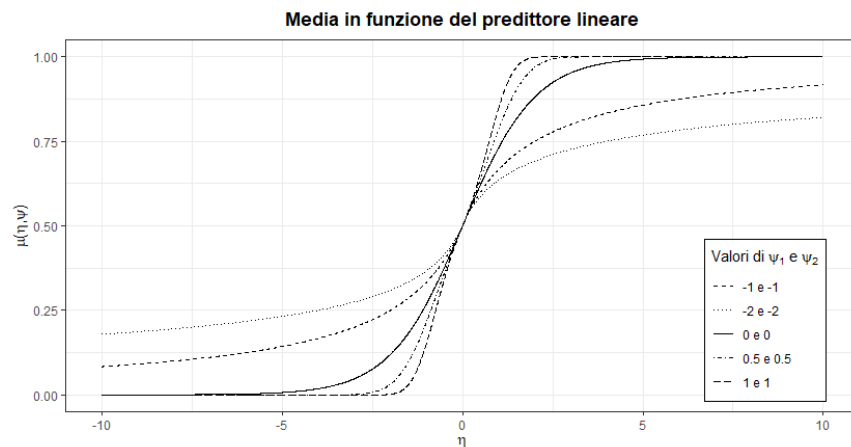


Figura 2.2.5: Media in funzione del predittore lineare per diversi valori di  $\psi_1$  e  $\psi_2$  con la famiglia di trasformazioni di Stukel per dati binari.



## 2.3 Quando usare un link non canonico

I GLM con link canonici garantiscono una semplice interpretazione del risultato, tuttavia non sempre forniscono il miglior adattamento ad un insieme di dati: è di interesse allora evitare errate specificazioni della funzione di legame.

Per poter comprendere quando la scelta di un link appartenente ad una classe di legami  $\mathcal{F}$  sia più vantaggiosa rispetto all'utilizzo del legame canonico, si possono adottare due approcci: il primo si basa sulla costruzione di un test d'ipotesi (Czado e Munk, 2000); il secondo si basa sull'utilizzo del fattore di Bayes (Raftery, 1996). Di seguito si farà riferimento al primo metodo.

### 2.3.1 Test d'ipotesi per GLM con link parametrici

Assumendo che il vero ed ignoto legame per un GLM con una data distribuzione della risposta sia membro della famiglia  $\mathcal{F}$ , al fine di evitare errate specificazioni di tale link si utilizza il seguente test d'ipotesi

$$\begin{cases} H_0 : \psi = \psi_* \\ H_1 : \psi \neq \psi_* \end{cases} \quad (2.22)$$

Se  $H_0$ , ovvero l'ipotesi nulla, non può essere rifiutata, ci si basa sul  $p$ -value fornito dalla statistica test come evidenza per poter mantenere la funzione di legame canonica nel modello (per la quale si ha  $\psi = \psi_*$ ). In caso contrario i parametri di regressione  $\beta$  e di legame  $\psi$  verranno stimati attraverso la massima verosimiglianza.

La questione è dunque capire quando è conveniente utilizzare una funzione di legame non canonica, e quanto grande deve essere l'evidenza per il modello canonico indicata dal  $p$ -value in (2.22).

Si considera ora la classe dei legami logistici generalizzati con trasformazione della funzione di link che ammette una modifica separata di entrambe le code (2.10, 2.11 e 2.12).

Si considera prima una misura di discrepanza  $\Delta$  tra i due legami utilizzati, e successivamente si considerano le curve di  $p$ -value generalizzate associate ad un test unilaterale per  $\Delta$ .

Per avere una migliore interpretazione del test d'ipotesi (2.22), quest'ultimo può essere riformulato nel seguente modo

$$\begin{cases} H_0 : \psi \notin [\psi_* - \psi_l, \psi_* + \psi_u] \\ H_1 : \psi \in [\psi_* - \psi_l, \psi_* + \psi_u], \end{cases} \quad (2.23)$$

dove  $\psi_l$  e  $\psi_u$  sono specifici valori maggiori di zero. Se  $H_0$  è rifiutata ad un dato livello di significatività  $\alpha$ , il legame canonico è valido con probabilità

d'errore  $\alpha$  in un intorno  $(\psi_* - \psi_l, \psi_* + \psi_u)$ .

Siano  $\lambda_{\min}(A)$  e  $\lambda_{\max}(A)$  gli autovalori più piccoli e più grandi, rispettivamente, di una matrice quadrata  $A$ . Sia  $I(\delta)$  la matrice d'informazione di Fisher,  $H(\delta)$  la matrice di informazione attesa,  $\delta_0(\beta_0, \psi_0)$  il vero valore del parametro  $\delta$  e  $s(\delta) = (s_1(\delta), \dots, s_{p+1}(\delta))$  il vettore degli *score*. Sono date le seguenti assunzioni:

- i.  $\lambda_{\min}(I_n) \rightarrow \infty$  per  $n \rightarrow \infty$ ;
- ii. esiste un intorno  $N$  di  $\delta_0$  tale che  $\lambda_{\min}(H_n(\delta)) \geq c[\lambda_{\max}(I_n)]^{1/2+\epsilon}$  con  $\delta \in N$ ,  $n \geq n_1$  con  $c, \epsilon > 0$  costanti e  $n_1$  un numero casuale;
- iii. si assume  $\{x_n, x \geq 1\} \subset K$  compatto e  $F(x_n^T \beta, \psi)$  due volte differenziabile in  $\beta$  e  $\psi$  e limitato per  $\{x_n, x \geq 1\} \subset K$  per valori fissati di  $\beta$  e  $\psi$ ;
- iv.  $n/\lambda_{\min}(I_n)$  uniformemente limitato per ogni  $n \geq 1$ .

Allora sotto le assunzioni (i), (iii) e (iv) per il sistema d'ipotesi (2.23), un livello  $\alpha$  per tale test è dato dalla regione di rifiuto

$$\mathcal{R} = \left\{ \hat{\psi} : \left| \phi \left( \frac{\hat{\psi} - \psi_* - \psi_u}{\hat{\sigma}_n(\psi, \beta)} \right) - \phi \left( \frac{-(\hat{\psi} - \psi_*) - \psi_l}{\hat{\sigma}_n(\psi, \beta)} \right) \right| \leq \alpha \right\}, \quad (2.24)$$

dove  $\hat{\sigma}_n^2$  è una stima consistente di  $\sigma_n^2(\psi, \beta)$ , ovvero dell'( $p+1, p+1$ )-esimo valore dell'inversa della matrice di informazione di Fisher  $I_n(\delta)$ . La regione d'accettazione per un test di livello  $1 - \alpha$  per il problema d'ipotesi (2.23) è pertanto data dalla regione complementare  $\mathcal{R}^C$ . Per applicare la (2.24) si deve prima stimare  $\sigma_n^2(\psi, \beta)$  con  $\hat{\sigma}_n^2 := \sigma_n^2(\hat{\psi}, \hat{\beta})$ , dove  $(\hat{\psi}, \hat{\beta})$  sono le stime di massima verosimiglianza per  $(\psi, \beta)$ .

I valori  $\psi_l$  e  $\psi_u$  sono fondamentali per poter ottenere la regione d'accettazione per l'ipotesi (2.23). Sia allora  $\hat{\beta}_\psi$  la stima di massima verosimiglianza per  $\beta$  con  $\psi$  fissato in un'arbitraria famiglia di legame  $\mathcal{F} = \{F(\cdot, \psi) : \psi \in \Psi\}$  e  $\eta_i(\hat{\beta}_\psi)$  il corrispondente  $i$ -esimo predittore.

Si determina una misura di discrepanza  $\Delta_{(\cdot)}$  dal modello canonico. Tale misura si ottiene calcolando la massima variazione nella stima della media della risposta quando si passa da un modello con legame canonico  $(\psi_*)$  ad un altro  $(\psi)$ ; allora

$$\Delta_{(m)}(\phi) = \max_{i=1, \dots, n} |F(\eta_i(\beta_\psi), \psi) - F(\eta_i(\beta_{\psi_*}), \psi_*)|, \quad (2.25)$$

dove  $\beta_\psi$  è il vero ignoto parametro di regressione in modello  $F(\cdot, \psi)$ . Tale risultato può essere utilizzato per il confronto di due qualsiasi famiglie con legami parametrici senza che uno di questi debba essere quello canonico, le conclusioni sono le medesime.

Successivamente si passa a definire il limite di tolleranza  $\Delta_0$ . Per il margine  $\Delta_0$  si assume che il modello canonico sia sufficientemente approssimato dal modello non canonico. Allora il problema d'ipotesi (2.23) è ora equivalente a

$$\begin{cases} \tilde{H}_0 : \Delta_m(\psi) > \Delta_0 \\ \tilde{H}_1 : \Delta_m(\psi) \leq \Delta_0, \end{cases} \quad (2.26)$$

dove le soglie  $\psi_* - \psi_l, \psi_* + \psi_u$  sono calcolate numericamente attraverso  $\Delta_m^{-1}(\Delta_0)$ .

Si nota che tale specificazione porta sempre ad avere due unici valori  $\psi_* - \psi_l^0 < \psi_* - \psi_u^0$  tali che  $\psi_* - \psi_l^0 < \psi < \psi_* - \psi_u^0$  in quanto il criterio  $\Delta_m(\psi)$  è strettamente unimodale con un unico valore di minimo in  $\psi_*$ . Allora  $\psi_l$  e  $\psi_u$  possono essere espressi in termini di  $\Delta_m$  in relazione anche al particolare modello scelto.

Vi sono altre misure di discrepanza che possono essere utilizzate e che non riguardano solo la variazione in termini assoluti della media della risposta d'interesse. Un'alternativa è fornita da

$$\Delta_r(\psi) = \max_{i=1, \dots, n} \begin{cases} \frac{F(\eta_i(\beta_\psi), \psi)}{F(\eta_i(\beta_{\psi_*}), \psi_*)} \text{ se } F(\eta_i(\beta_\psi), \psi) < F(\eta_i(\beta_{\psi_*}), \psi_*), \\ \frac{F(\eta_i(\beta_{\psi_*}), \psi_*)}{F(\eta_i(\beta_\psi), \psi)} \text{ altrimenti.} \end{cases} \quad (2.27)$$

Tale formulazione per la misura di discrepanza trova applicazione soprattutto nel caso binomiale, dove gli effetti delle covariate vengono interpretati utilizzando il rapporto delle quote. Si considera, ad esempio, la variazione massima nella quota che può essere stimata con

$$\Delta_o(\psi) = \max_{i=1, \dots, n} \begin{cases} o_i(\psi) \text{ se } F(\eta_i(\beta_\psi), \psi) > F(\eta_i(\beta_{\psi_*}), \psi_*), \\ 1/o_i(\psi) \text{ altrimenti,} \end{cases} \quad (2.28)$$

dove

$$o_i(\psi) = \frac{F(\eta_i(\beta_\psi), \psi)}{1 - F(\eta_i(\beta_\psi), \psi)} \bigg/ \frac{F(\eta_i(\beta_{\psi_*}), \psi_*)}{1 - F(\eta_i(\beta_{\psi_*}), \psi_*)}.$$

Il test d'ipotesi (2.22) per ipotesi nulle semplici non può essere utilizzato come una misura di evidenza; il passaggio al test d'ipotesi (2.23) non risolve totalmente tale questione ma ammette la specificazione di ipotesi più flessibili per far fronte al problema della non corretta specificazione della *direzione* del rifiuto (contro o a favore del modello canonico). Inoltre, grazie a quest'ultima formulazione, la direzione può essere espressa e quantificata anche in termini di entità di interesse (come la quota). Si vuol capire quale, tra errore di primo e secondo tipo, ora è di primario interesse, e, a tale scopo, si utilizzano le curve *p-value*.

In aggiunta ad un test d'ipotesi con limiti prefissati  $\Delta_0$  si considera allora la funzione  $\hat{P}$  data da

$$\hat{P}(C, \hat{\sigma}_n^2, \psi_l, \psi_u) = \psi\left(\frac{C - \theta_1}{\hat{\sigma}^2}\right) - \psi\left(\frac{-(C + \theta_1)}{\hat{\sigma}^2}\right) = \alpha, \quad (2.29)$$

dove  $\theta_1 = \frac{\psi_u + \psi_l}{2}$  e  $C > 0$ .

Si indica con  $\hat{P}$  la curva  $p$ -value associata al test (2.23) per discriminare tra le ipotesi. Per un dato insieme di osservazioni si ha che  $\hat{P}(\hat{\psi}_n, \hat{\sigma}_n^2, \psi_l, \psi_u)$  è una superficie bidimensionale dove il livello  $\alpha = \hat{P}$  fornisce i limiti minimi  $\psi_* - \psi_l$  e  $\psi_* + \psi_u$  entro i quali l'ipotesi nulla  $H_0$  può essere rifiutata al livello  $\alpha$ . In particolare, quando  $\psi_* - \psi_l = \psi_u = \psi_* + \psi_*$ ,  $1 - \hat{P}$  rappresenta il classico  $p$ -value fornito dal test (2.23). E' di interesse per una valida interpretazione dell'approccio appena proposto, che la superficie in due dimensioni  $\hat{P}$  venga ridotta ad una curva unidimensionale in riferimento ad una particolare misura di discrepanza.

Scelto allora un criterio tra  $\Delta_o\psi$ ,  $\Delta_m\psi$  o  $\Delta_r\psi$ , tale superficie dipende da  $\Delta_{(\cdot)}$  solo dalla seguente relazione:  $\hat{P}(\cdot, \cdot, \Delta_{(\cdot)}) = \hat{P}(\cdot, \cdot, \psi_l, \psi_u)$ .

Dunque le curve  $p$ -value per  $\Delta_{(\cdot)}$  rappresentano un'utile misura d'evidenza all'interno di un test d'ipotesi.

### 2.3.2 Procedura per l'identificazione del link corretto

Di seguito si esamina una procedura per valutare l'adeguatezza di un link scelto attraverso l'introduzione di un insieme di variabili esplicative. Sia  $g_0(\mu)$  la funzione di legame ipotizzata per un modello e sia  $g_*(\mu)$  la funzione di legame corretta ed ignota per il medesimo modello. Seguendo il ragionamento illustrato nei paragrafi precedenti, si considera il link ipotizzato all'interno di una famiglia di funzioni di legame, con  $\alpha$  e  $\delta$  due parametri di legame.

Si considera allora

$$\text{Link ipotizzato: } g_0(\mu) = g(\mu, \alpha_0, \delta_0), \quad (2.30)$$

$$\text{Link corretto: } g_*(\mu) = g(\mu, \alpha_*, \delta_*). \quad (2.31)$$

Utilizzando l'approssimazione al primo ordine di Taylor per il legame ipotizzato, si ottiene

$$g_*(\mu) \simeq g_0(\mu) + (\alpha_* - \alpha_0)D_\alpha(g_0) + (\delta_* - \delta_0)D_\delta(g_0), \quad (2.32)$$

dove

$$D_\alpha(g_0) = \left\{ \frac{\partial}{\partial \alpha} g(\mu, \alpha, \delta) \right\} \Big|_{\alpha=\alpha_0, \delta=\delta_0},$$

e si ha un'espressione analoga per  $D_\delta(g_0)$ .

Si approssima ora il legame corretto  $g_*(\mu) = \mathbf{x}\beta$  nel seguente modo

$$g_0(\mu) = \mathbf{x}\beta + \mathbf{z}\gamma, \quad (2.33)$$

con  $\mathbf{x} = (D_\alpha(g_0), D_\delta(g_0))$  e  $\gamma^T = -(\alpha_* - \alpha_0, \delta_* - \delta_0)$ .

In tal modo il problema della modifica del legame è stato riformulato in termini del link ipotizzato aggiungendo dei fattori alla componente sistematica che descrivono le differenze tra le due funzioni di legame.

Il vantaggio dato da tale formulazione sta nel fatto che ora si può considerare una procedura in due passi:

1. dapprima si adatta il modello con funzione di legame  $g_0(\mu) = \mathbf{x}\beta$ : tale adattamento porterà ad avere una stima  $\hat{\beta}$  per i coefficienti di regressione con i conseguenti valori previsti  $g_0(\hat{\mu}) = \mathbf{x}\hat{\beta}$ , dai quali si ricava  $\hat{z} = z_{|\beta=\hat{\beta}}$ ;
2. si testa allora l'adeguatezza del legame riformulando il modello utilizzando  $\hat{z}$  come ulteriore insieme di variabili esplicative.

Una riduzione della devianza indicherà un allontanamento dalla funzione di legame ipotizzata.

Si consideri ad esempio il caso della regressione binomiale con funzione di legame alternativa alla logistica data dalla (2.15), dove ora  $\alpha_* = \psi_1$  e  $\delta_* = \psi_2$ . L'adattamento di tale modello porta ad avere  $\hat{\beta}$  e  $g_0(\hat{\mu})$ . Il legame logistico si ha per  $\alpha_0 \rightarrow 0$  e  $\delta_0 \rightarrow 0$ , in modo tale che  $\gamma^T = -(\alpha_*, \delta_*)$ . Allora  $\mathbf{z}^T = (\frac{1}{2}(\log^2(\hat{\theta}) - \log^2(1 - \hat{\theta})), -\frac{1}{2}(\log^2(\hat{\theta}) + \log^2(1 - \hat{\theta})))$ .

## 2.4 Conclusioni

Le funzioni di legame parametriche possono apportare un miglioramento nell'adattamento dei GLM: la flessibilità che queste introducono, infatti, consente di applicare i modelli ad un più ampio insieme di casi.

La costruzione di link parametrici inizia dalla specificazione di famiglie di legami standardizzate. L'ortogonalizzazione dei parametri  $\beta$  e  $\psi$  ha un ruolo fondamentale dal momento che consente di ridurre la varianza dei parametri di regressione, causata dall'introduzione della stima dei parametri di link, e porta ad avere stime numericamente stabili (e quindi anche meglio interpretabili) per le stime  $\hat{\beta}$  corrispondenti a diversi parametri di legame.

Le modifiche alla funzione di legame possono essere simmetriche o meno, e, attraverso opportune determinazioni dei parametri di link, possono riportare al legame canonico per un dato modello. È di interesse allora capire, risolvendo opportuni test d'ipotesi, quando tale modifica risulta necessaria. In particolare si pone l'attenzione sul termine di guadagno che l'applicazione

di tali trasformazioni apporta. Tale guadagno lo si misura in termini di devianza residua.

Nel prossimo capitolo si porrà l'attenzione sulla stima dei parametri di modelli con link parametrici servendosi anche di applicazioni a dati reali.

## Capitolo 3

# Stima dei parametri di link

Si pone ora l'attenzione sulla stima del parametro di legame e dei parametri di regressione nei GLM con funzioni di legame parametriche. Dal momento che si studiano due insiemi di parametri congiuntamente si farà ricorso alla verosimiglianza profilo. Verranno proposti alcuni esempi di stima ed applicazioni a dati reali delle trasformazioni trattate.

Per le analisi effettuate è stato utilizzato il software R (versione 3.5.2, [www.r-project.org/](http://www.r-project.org/)).

### 3.1 Teoria della stima del parametro di legame

Nei GLM con funzioni di legame parametriche definiti come nel Paragrafo 2.1, il parametro di regressione  $\beta$  ed il parametro di legame  $\psi$  sono stimati congiuntamente attraverso massima verosimiglianza. Più precisamente, se il vero legame  $F$  è un membro della famiglia di legame  $\mathcal{F}$ , sotto condizioni di regolarità, la stima di massima verosimiglianza congiunta per il parametro  $\delta = (\psi, \beta)$ ,  $\hat{\delta} = (\hat{\psi}, \hat{\beta})$ , è consistente. Il parametro  $\phi$  ignoto viene invece stimato da un opportuno stimatore ottenuto con il metodo dei momenti (1.32).

Come per i modelli lineari generalizzati ordinari, sapendo che  $\mu_i = b'(\theta_i)$  e  $\mu_i = F(\eta_i(\beta), \psi)$  e considerando ora  $a_i(\phi) = a(\phi) \forall i$ , la log-verosimiglianza  $\ell(\delta)$  può esser scritta come

$$\ell(\delta) = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^n c(y_i, \phi). \quad (3.1)$$

Gli elementi della funzione punteggio, sapendo che  $\mu_i = F(\eta_i(\beta), \psi)$ , sono

$$\begin{aligned} \frac{\partial \mu_i}{\partial \psi} &= F_{i2}, \quad i = 1, \dots, n, \\ \frac{\partial \mu_i}{\partial \beta_j} &= x_{ij} F_{i1}, \quad j = 1, \dots, p, \quad i = 1, \dots, n. \end{aligned} \quad (3.2)$$

In particolare si ha  $F_{i1} = \frac{\partial}{\partial \eta} F(\eta, \psi)|_{\eta=\eta_i}$  e  $F_{i2} = \frac{\partial}{\partial \psi} F(\eta, \psi)|_{\eta=\eta_i}$  per  $i = 1, \dots, n$ . Sia  $d_i = \frac{d\theta_i}{d\mu_i}$ , allora gli elementi della funzione punteggio sono

$$s_j(\delta) = \frac{\partial}{\partial \beta_j} \ell(\delta) = \sum_{i=1}^n \frac{d\theta_i \partial \mu_i}{d\mu_i \partial \beta_j} \left[ \frac{y_i - \mu_i}{a(\phi)} \right] = a(\phi)^{-1} \sum_{i=1}^n d_i x_{ij} F_{i1}(y_i - \mu_i), \quad (3.3)$$

$$s_{p+1}(\delta) = \frac{\partial}{\partial \psi} \ell(\delta) = \sum_{i=1}^n \frac{d\theta_i \partial \mu_i}{d\mu_i \partial \psi} \left[ \frac{y_i - \mu_i}{a(\phi)} \right] = a(\phi)^{-1} \sum_{i=1}^n d_i x_{ij} F_{i2}(y_i - \mu_i). \quad (3.4)$$

L'informazione di Fisher  $I(\delta)$  è espressa nel seguente modo

$$I(\delta) = a(\phi)^{-1} \begin{pmatrix} I_{\beta, \beta} & I'_{\beta, \psi} \\ I_{\beta, \psi} & I'_{\psi, \psi} \end{pmatrix}, \quad (3.5)$$

dove  $I_{\beta, \beta}$  è una matrice di dimensioni  $(p+1) \times (p+1)$  con elementi

$$(I_{\beta, \beta})_{rs} = \sum_{i=1}^n x_{is} x_{ir} F_{i1}^2 d_i, \quad r \geq 0, \quad s \geq 0,$$

$I_{\beta, \psi}$  è un vettore  $(p+1)$  con

$$(I_{\beta, \psi})_r = \sum_{i=1}^n x_{ir} F_{i1} F_{i2} d_i, \quad r \geq 0,$$

e  $I_{\psi, \psi}$  è un numero reale dato da

$$I_{\psi, \psi} = \sum_{i=1}^n F_{i2}^2 d_i.$$

Per trovare la stima di massima verosimiglianza  $\hat{\delta}$  si devono porre le due equazioni (3.3) e (3.4) uguali a zero. Se si pone la (3.3) a zero fissando il parametro  $\psi$  per un dato legame, questo equivale a trovare la stima di massima verosimiglianza di  $\beta$ ,  $\hat{\beta}$ , come nei GLM ordinari. Allora un modo per trovare  $\hat{\delta}$  consiste nel ricorrere alla verosimiglianza profilo<sup>1</sup>, calcolando prima  $\hat{\beta}_\psi$ , la stima di massima verosimiglianza per un GLM con link fissato e successivamente estrarre il valore  $\hat{\psi}$  che massimizza la log-verosimiglianza per il legame scelto.

Si consideri ora la modifica della funzione di legame come in (2.20) e (2.21). Il caso per cui  $h(\eta) = \eta$  lo si ha per  $\psi_1 = \psi_2 = 0$ , e corrisponde al

<sup>1</sup>In un modello statistico parametrico con funzione di verosimiglianza  $L(\delta)$ , dove  $\delta = (\beta, \psi)$ , si dice verosimiglianza profilo per  $\beta$ , con  $\beta$  parametro di interesse, la funzione  $L_p(\beta) = L(\beta, \hat{\psi}_\beta)$ , dove  $\hat{\psi}_\beta$  è la stima di massima verosimiglianza di  $\psi$  nel sottomodello con  $\beta$  fissato.



legame logistico. Si può ricorrere alla stima di massima verosimiglianza per i parametri lineari utilizzando una procedura iterativa con i minimi quadrati pesati iterati che utilizza la statistica log-rapporto di verosimiglianza (Brillinger *et al.*, 1985).

Tale procedimento per la stima può esser così riassunto. Sia  $\beta$  un vettore di parametri lineari di regressione, e  $\psi$  un vettore a due componenti di parametri non lineari. Per fissati valori di  $\psi$  si ha uno specifico GLM. La stima di massima verosimiglianza per  $\beta$  per un dato  $\psi$  è  $\hat{\beta}_\psi$ . Ad ogni passo viene dato un valore iniziale per  $\beta$  corrispondente alla stima di massima verosimiglianza raggiunta al passaggio precedente, iniziando da  $\hat{\beta}_0$ , la stima di massima verosimiglianza sotto il modello logistico. La coppia  $(\hat{\beta}, \hat{\psi})$  che minimizza la devianza rappresenta allora la stima di massima verosimiglianza.

Il software maggiormente utilizzato per l'applicazione di questo algoritmo è GLIM (Aitkin, 2018). Tale algoritmo permette di ciclare tra i due gruppi di parametri massimizzando di volta in volta la log-verosimiglianza condizionata ad un set di parametri mantenendo gli altri fissati. Ad ogni passo viene posta la funzione score uguale a zero e viene risolta l'equazione. Per un  $\psi$  fissato si ha il procedimento precedentemente descritto che porta ad avere  $\hat{\beta}_\psi$ , mentre per un fissato  $\beta$  si hanno funzioni non lineari in  $\psi$  che possono esser risolte ricorrendo all'algoritmo di Newton-Raphson. Tuttavia va tenuto presente che tale procedimento basato sulla log-verosimiglianza richiede algoritmi complessi e talvolta può portare alla divergenza delle stime.

Per trovare una stima della matrice di varianza asintotica per  $(\hat{\beta}, \hat{\psi})$ ,  $V_{\hat{\beta}, \hat{\psi}}$ , si inverte la matrice di informazione attesa di Fisher. Per  $\psi = (\psi_1, \psi_2)$  si ha allora

$$V_{\hat{\beta}, \hat{\psi}} = \begin{pmatrix} X^T W X & X^T C \\ C^T X & U \end{pmatrix}^{-1} = \begin{pmatrix} B^{-1} + M A^{-1} M^T & -M A^{-1} \\ -A^{-1} M^T & A^{-1} \end{pmatrix}, \quad (3.6)$$

dove  $B^{-1} = (X^T W X)^{-1}$  è la matrice di varianza-covarianza di  $\hat{\beta}$  dato  $\hat{\psi}$ .  $X$  è la matrice di disegno del modello,  $W$  è la matrice dei pesi della procedura dei minimi quadrati pesati.  $M = B^{-1} X^T C$  è una matrice  $p \times 2$  con  $p$  numero di variabili nel modello, e  $A^{-1} = (U - C^T X B^{-1} X C)^{-1}$  è la matrice di varianza per  $\hat{\psi}$ , data da

$$\begin{pmatrix} \mu(1 - \mu) \frac{\partial}{\partial \eta} h_{\psi_1}(\eta) h_{\psi}(\eta) \\ \mu(1 - \mu) \frac{\partial}{\partial \eta} h_{\psi_2}(\eta) h_{\psi}(\eta) \end{pmatrix}, \quad (3.7)$$

con  $U = \text{diag}(u_j) = \text{diag}\left(\sum_{i \in G_j} \mu_i(1 - \mu_i) \left(\frac{\partial}{\partial \psi_j} h_{\psi}(\eta_i)\right)\right)$ , con  $j = 1, 2$ .

$G_1$  è l'insieme dei punti per i quali si ha  $\eta > 0$ , mentre  $G_2$  quello per valori di  $\eta$  negativi.

Si consideri ora un problema di inferenza parziale, dove il parametro di interesse è  $\psi$  mentre  $\beta$  è un parametro di disturbo. Un modo per ricercare una regione entro la quale una modifica della funzione di legame è accettabile, consiste nel ricorrere al log-rapporto di verosimiglianza parziale.

### 3.1.1 Un esempio

Volendo ora esemplificare il procedimento per discutere un test d'ipotesi viene presa a riferimento la funzione di legame modificata sulla sola coda sinistra di Stukel (2.21). Il test è il seguente

$$\begin{cases} H_0 : \psi = \psi_0 = 0 \\ H_1 : \psi \neq \psi_0. \end{cases} \quad (3.8)$$

La statistica test del log-rapporto di verosimiglianza profilo per il test d'ipotesi che pone sotto  $H_0$  il modello logistico e sotto  $H_1$  il modello con funzione di legame modificata, è la seguente

$$W_p^{H_0} = 2 \left( \ell(\hat{\eta}, \hat{\psi}) - \ell(\hat{\eta}_0, \psi_0) \right), \quad (3.9)$$

dove  $\hat{\eta}_0$  rappresenta la stima di massima verosimiglianza di  $\eta$  per il modello sotto  $H_0$ .

Nel caso in questione sotto  $H_1$  si ha

$$h_\psi(\eta) = \log(\mu/(1 - \mu)),$$

e quindi

$$\mu_\psi = \frac{\exp\{h_\psi(\eta)\}}{1 + \exp\{h_\psi(\eta)\}}, \quad (3.10)$$

mentre sotto l'ipotesi nulla si ha il legame logistico dato da

$$\mu = \frac{\exp\{\eta\}}{1 + \exp\{\eta\}}. \quad (3.11)$$

Allora la statistica test può esser espressa nel seguente modo:

$$\begin{aligned} & 2 \left( \ell(\hat{\eta}, \hat{\psi}) - \ell(\hat{\eta}_0, \psi_0) \right) = \\ & = 2 \left[ \sum_i \left[ y_i \log \frac{\exp\{h_\psi(\eta_i)\}}{1 + \exp\{h_\psi(\eta_i)\}} + (n - y_i) \log \left( 1 - \frac{\exp\{h_\psi(\eta_i)\}}{1 + \exp\{h_\psi(\eta_i)\}} \right) \right] \right. \\ & \quad \left. - \sum_i \left[ y_i \log \frac{\exp\{\eta_i\}}{1 + \exp\{\eta_i\}} + (n - y_i) \log \left( 1 - \frac{\exp\{\eta_i\}}{1 + \exp\{\eta_i\}} \right) \right] \right], \end{aligned} \quad (3.12)$$

e dunque sostituendo le (3.10) e (3.11)

$$\sum_i 2 \left[ y_i \log \frac{\mu_\psi}{\mu} + (n - y_i) \log \frac{1 - \mu_\psi}{1 - \mu} \right] = W_p(\psi). \quad (3.13)$$

Una regione di accettazione per tale ipotesi è definita da

$$\hat{\Psi}(y) : \left\{ \psi : W_p(\psi) \leq \chi_{1-\alpha;1}^2 \right\}, \quad (3.14)$$

dove  $\alpha$  è il livello di significatività fissato per il test d'ipotesi (3.8).

Per poter effettuare test con ipotesi alternativa unilaterale si può ricorrere alla radice col segno della statistica  $W_p(\psi_0)$  definita come

$$r_p(\psi_0) = \text{sgn}(\hat{\psi} - \psi_0) \sqrt{W_p(\psi_0)},$$

la quale ha distribuzione asintotica normale standard sotto  $H_0$ , ossia

$$r_p(\psi_0) \sim N(0, 1). \quad (3.15)$$

### 3.2 Stima del parametro di legame con LikelihoodAsy

In tale paragrafo si studia un problema di verosimiglianza profilo attraverso l'utilizzo della libreria *likelihoodAsy* di R (Bellio e Pierce, 2018), la quale consente di trovare le statistiche  $r_p$  e  $r_p^*$ . Con  $r_p^*$  s'intende un miglioramento asintotico di  $r_p$ , la cui distribuzione nulla è  $N(0, 1)$  al terzo ordine (si veda ad esempio Brazzale *et al.*, 2007).

Si consideri il dataset *babies* contenuto nel pacchetto *cond* di R. Il dataset contiene informazioni riguardanti il numero di bambini che in un dato giorno, indicato da un'opportuna variabile, hanno e non hanno pianto, insieme ad una variabile dicotomica per indicare se è stato svolto o meno un trattamento. La soglia di significatività fissata per tale analisi è del 5%.

Indicando con  $Y$  la variabile risposta il modello è così formulato

$$Y_i \sim Bi(m_i, \mu_i), \quad i = 1, \dots, 36,$$

dove  $m_i$  indica la numerosità di ogni gruppo dato dalla combinazione tra giorno e trattamento.

La funzione di legame presa a riferimento è la funzione appartenente alla famiglia parametrica di Burr-Prentice, per la quale la media della risposta è così definita

$$\mu = 1 - \frac{1}{(1 + \exp(\eta))^\psi}, \quad \text{per } \eta \in \mathbb{R} \text{ e } \psi > 0. \quad (3.16)$$

È stata utilizzata la funzione *rstar* della libreria *likelihoodAsy*. Come valori iniziali per i coefficienti di regressione  $\beta$  sono state utilizzate le stime di massima verosimiglianza date dall'adattamento del GLM con funzione di legame canonica (logit). Il valore che si è scelto di porre sotto ipotesi nulla per  $\psi$  è 1, in quanto tale valore restituisce la funzione di legame logistica. La stima di massima verosimiglianza ottenuta per il parametro di legame è:  $\hat{\psi} = 0.530$ , con standard error di 0.864.

Il valore  $r_p$  ottenuto è di  $-0.365$  con  $p$ -value pari a 0.357; il valore della statistica alla Wald è invece  $-0.544$  con  $p$ -value 0.293. Il valore di  $r_p^*$  che si ottiene è  $-0.383$ . Dunque all'interno di un test d'ipotesi

$$\begin{cases} H_0 : \psi = 1 \\ H_1 : \psi < 1, \end{cases} \quad (3.17)$$

dove sotto l'ipotesi nulla si è posto il modello logistico, tale procedura porta ad accettare  $H_0$  ad un livello di significatività del 5%, con un  $p$ -value pari a 0.351.

I grafici della verosimiglianza profilo e della devianza per il parametro di interesse  $\psi$  fatto variare all'interno di una griglia di valori sono mostrati in Figura (3.2.1) e (3.2.2).

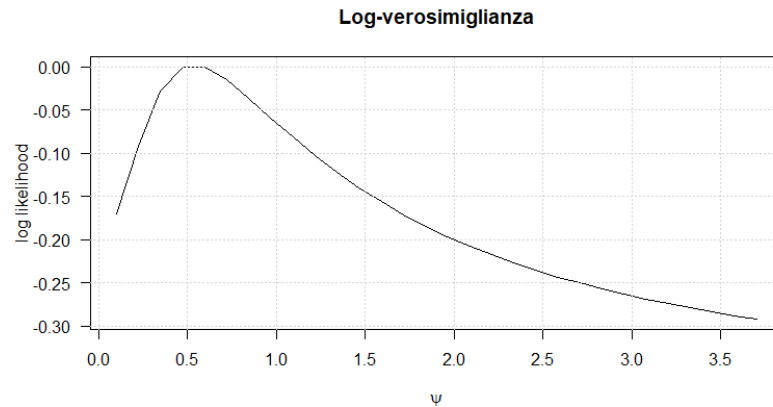


Figura 3.2.1: Log-verosimiglianza profilo con funzione parametrica di Burr-Prentice al variare del parametro  $\psi$ .

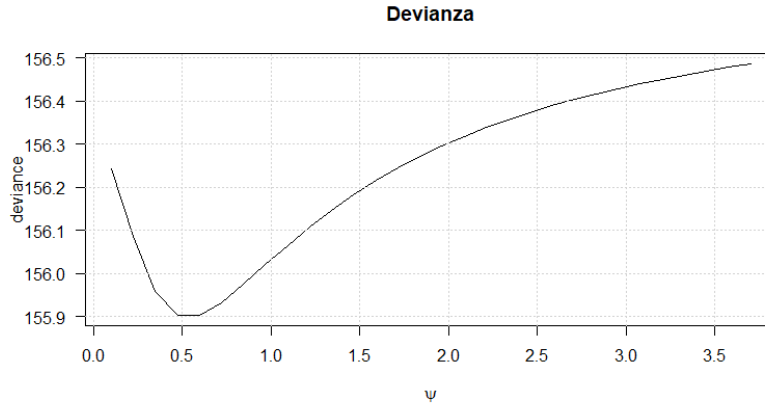


Figura 3.2.2: Devianza per il confronto tra il modello con funzione parametrica di Burr-Prentice con il modello logistico al variare del parametro  $\psi$ .

### 3.3 Modifica delle code della funzione di legame

Le analisi riportate di seguito sono state condotte con l'utilizzo una funzione che consente di modificare il legame logistico come in (2.12) portando così ad una *regressione logistica con modifica di una coda* (Flach, 2014).

#### 3.3.1 Regressione logistica modificata

Per le analisi riportate di seguito si utilizza il dataset *downs.bc* disponibile nel pacchetto *boot* di R. I dati si riferiscono ad uno studio su larga scala effettuato negli anni '60 da parte del *British Columbia Health Surveillance Registry* per investigare la relazione tra l'età della madre e l'incidenza della sindrome di Down. La soglia di significatività fissata per le analisi è del 5%.

Verranno utilizzate le funzioni *glmProfile* e *psi2LOGIT* (Flach, 2014). La prima funzione consente di studiare il valore di  $\psi_2$  (per un dato  $\eta_0$ ) a cui è associato il modello che meglio si adatta ai dati. Lo scopo allora è scegliere il valore di  $\psi_2$  al quale è associato il modello con minore devianza residua. La seconda funzione, invece, definisce la funzione logit con coda sinistra modificata.

Sia  $Y$  la variabile risposta. Il modello inizialmente formulato è

$$Y_i \sim Bi(m_i, \mu_i), \quad i = 1, \dots, 30,$$

dove  $m_i$  rappresenta la numerosità dei gruppi ed ogni gruppo conta il numero di nati da madri di una fissata età.

Si è dapprima adattato un modello di regressione logistica senza aver applicato alcuna modifica alla funzione di legame come specificato nel seguente modo

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = x_i\beta = \eta_i, \quad i = 1, \dots, 30,$$

con  $x_i$  età della madre riferita al gruppo  $i$ -esimo. L'output è riportato in Tabella (3.3.1) .

Tabella 3.3.1: Output per i coefficienti del modello di regressione logistica per il dataset *Downs.bc*.

<b>Coefficienti</b>	Stima	Std. Error	z value	<i>p</i> - value
Intercetta	-10.56	0.21	-49.25	< 0.001
età	0.14	0.01	21.25	< 0.001

La devianza residua è 184.03 su 28 gradi di libertà, l'indice di Akaike è invece  $AIC = 326.91$ .

Si osserva che applicare la trasformazione di tale coda alla funzione logit ponendo  $\psi_2 = 1$  porta ad ottenere il medesimo risultato appena presentato.

Si considera ora la modifica della coda sinistra della funzione logit. Per poter applicare una trasformazione a tale funzione si sono dovuti specificare in un primo momento le componenti che hanno permesso la creazione della funzione di legame modificata. In particolare le specificazioni necessarie sono state le seguenti:

- **la funzione di legame:**  $g(\mu, \psi)$ : da  $\eta = g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$  ora si ha

$$\begin{aligned} h_{\eta_0}(\eta, \psi) &= \log\left(\frac{\mu}{1-\mu}\right) \\ \Leftrightarrow \eta &= (h_{\eta_0}(\eta, \psi))^{-1} \left( \log\left(\frac{\mu}{1-\mu}\right) \right) \\ \Rightarrow g(\mu, \psi) &= (h_{\eta_0}(\eta, \psi))^{-1} \left( \log\left(\frac{\mu}{1-\mu}\right) \right); \end{aligned}$$

- **la funzione inversa della funzione di legame:**  $F(\eta, \psi)$ :

$$\begin{aligned} F(\eta, \psi) &= \frac{\exp\{h_{\eta_0}(\eta, \psi)\}}{1 + \exp\{h_{\eta_0}(\eta, \psi)\}} = \\ &= \frac{1 + \exp\{h_{\eta_0}(\eta, \psi)\}}{1 + \exp\{h_{\eta_0}(\eta, \psi)\}} - \frac{1}{1 + \exp\{h_{\eta_0}(\eta, \psi)\}} = \\ &= 1 - \frac{1}{1 + \exp\{h_{\eta_0}(\eta, \psi)\}}; \end{aligned}$$

- **la derivata rispetto al predittore della funzione inversa:**  $\frac{\partial \mu}{\partial \eta} = \frac{\partial}{\partial \eta} F(\eta, \psi)$ :

$$\begin{aligned} \frac{\partial}{\partial \eta} F(\eta, \psi) &= \frac{\partial}{\partial \eta} \left( \frac{\exp\{h_{\eta_0}(\eta, \psi)\}}{1 + \exp\{h_{\eta_0}(\eta, \psi)\}} \right) = \frac{\left( \frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \psi) \right) \exp\{h_{\eta_0}(\eta, \psi)\}}{(1 + \exp\{h_{\eta_0}(\eta, \psi)\})^2} = \\ &= \left( \frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \psi) \right) F(\eta, \psi) (1 - F(\eta, \psi)). \end{aligned}$$

Sia  $h_{\eta_0}(\eta, \psi = \psi_2)$  la modifica alla coda sinistra della funzione di legame logit come in (2.12). Il modello ora specificato ha funzione di legame

$$g(\mu_i, \psi_2) = (h_{\eta_0}(\eta, \psi))^{-1} \left( \log \left( \frac{\mu}{1 - \mu} \right) \right) \quad i = 1, \dots, 30,$$

con  $x_i$  età della madre riferita al gruppo  $i$ -esimo, dove

$$\eta = (h_{\eta_0}(\eta, \psi))^{-1}(y) = \begin{cases} y & \text{se } \eta \geq \eta_0 \\ \eta_0 + 1 - \exp(\eta_0 - y) & \text{se } \eta < \eta_0 \text{ e } \psi_2 = 0 \\ \eta_0 + 1 - (\psi_2(\eta_0 - y) + 1)^{\frac{1}{\psi_2}} & \text{se } \eta < \eta_0 \text{ e } \psi_2 \neq 0. \end{cases} \quad (3.18)$$

Prima di procedere all'applicazione della trasformazione è bene ricercare i valori di  $\psi_2$  e  $\eta_0$  che portano ad avere una maggior riduzione della devianza residua. Si pone inizialmente  $\eta_0 = 0$  e si applica la funzione *glmProfile*. Tale funzione riporta i valori della devianza residua per  $\psi_2$  fatto variare in un dato range e l'intervallo di confidenza al 95% per il valore del parametro associato alla minima devianza. Tale intervallo è stato calcolato aggiungendo 2 al minimo della devianza e prendendo i valori di  $\psi$  corrispondenti a tali altezze (Flach, 2014).

La devianza residua minima di 102.44 si ha per  $\psi_2 = 0.17$ . Un intervallo di confidenza per  $\psi_2$  al 95% è (0.135, 0.215).

Tale intervallo dà un'indicazione dell'adeguatezza della trasformazione alla funzione di legame: se infatti contiene il valore  $\psi = 1$ , ricordando che si sta utilizzando la trasformazione data da (2.12), allora la modifica non porta alcun miglioramento (per  $\psi = 1$  infatti si torna ad avere il legame logistico), e non vi è un importante abbassamento della devianza residua. Se invece l'intervallo di confidenza non contiene l'uno la modifica della funzione di legame sembra appropriata. Allora adattando il modello con  $\psi_2 = 0.17$  e  $\eta_0 = 0$  si ottiene il risultato riportato in Tabella 3.3.2:

Come si nota dalla Figura 3.3.1, la devianza residua si è abbassata a 102.44 e l'indice di Akaike è ora 245.32. Vi è un netto miglioramento nell'adattamento del modello ai dati.

Ponendo  $\eta_0 = 0.5$  si ottiene un lieve miglioramento: la devianza residua si

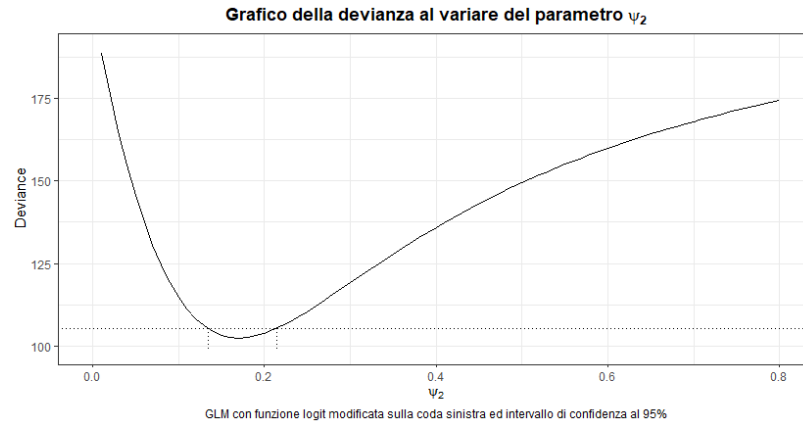


Figura 3.3.1: Devianza per il modello di regressione logistica per vari valori di  $\psi_2$  e con  $\eta_0 = 0$ .

Tabella 3.3.2: Output per i coefficienti di un modello di regressione logistica.

Coefficienti	Stima	Std. Error	z value	<i>p</i> – value
Intercetta	–212.77	6.07	–35.08	< 0.001
età	4.35	0.14	30.17	< 0.001

abbassa a 102.24 con  $AIC = 245.12$ . Tuttavia la differenza non è apprezzabile e si ritiene opportuno proseguire le analisi assegnando con  $\eta_0 = 0$ .

I risultati che si ottengono sono chiaramente visibili: la modifica della coda sinistra incide profondamente sulla bontà d’adattamento del modello ai dati. A seguito dell’adattamento del modello logistico modificato si può misurare in termini percentuali il miglioramento della devianza che si è ottenuto con

$$\text{miglioramento \%} = \left| \frac{\text{devianza del modello modificato}}{\text{devianza del modello ordinale}} - 1 \right| \cdot 100\%.$$

Per il caso presentato tale miglioramento è pari a 44.34%. Dunque si ha un adattamento del modello modificato ai dati del 44.34% migliore a quello che si ha con l’utilizzo del modello logistico.

### 3.3.2 Confronto di due modifiche della funzione logistica

Si considerano ora le funzioni *glmProfile*, *psi1LOGIT* e *psi2LOGIT* (Flach, 2014) per la ricerca dei valori ottimali per i parametri di legame  $\psi_1$  e  $\psi_2$  per la modifica della coda destra e sinistra della funzione logistica separatamente. Per le analisi effettuate di seguito si considera il dataset *Beetles* (Bliss, 1935). Tale dataset riporta il numero di scarafaggi esposti e di scarafaggi morti a seguito di cinque ore di esposizione a varie dosi di un’insetticida, il



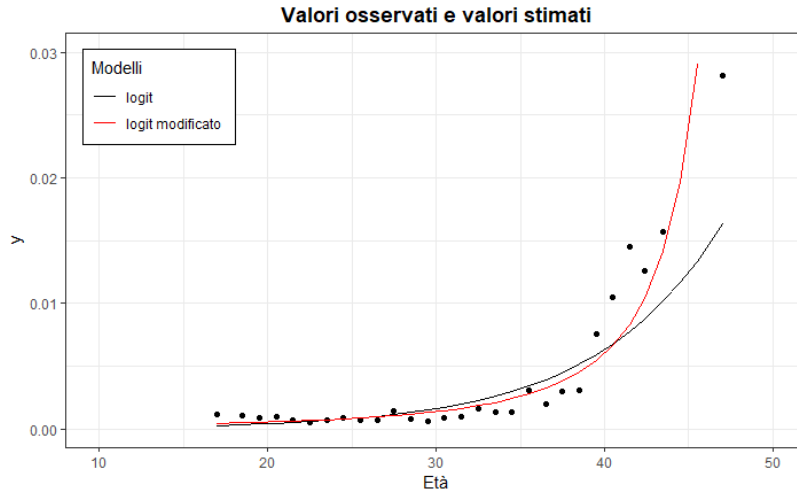


Figura 3.3.2: Modello logistico e modello logistico con coda modificata con  $\psi_2 = 0.17$  e  $\eta_0 = 0$ .

solfuro di carbonio ( $CS_2$ ). Sia  $Y$  la variabile risposta. Il modello inizialmente ipotizzato è così formulato

$$Y_i \sim Bi(m_i, \pi_i), \quad i = 1, \dots, 8,$$

con  $m_i$  numerosità di ogni singolo gruppo di insetti esposti, dove i gruppi sono individuati dalla concentrazione dell'insetticida.

L'adattamento del modello con funzione di legame logistica ha generato il seguente risultato riportato in Tabella 3.3.3.

Tabella 3.3.3: Output per i coefficienti del modello di regressione logistica per il dataset *Beetles*.

Coefficienti	Stima	Std. Error	z value	p-value
Intercetta	-60.72	5.18	-11.72	< 0.001
dose	34.27	2.91	11.77	< 0.001

La devianza residua è 11.23, su 6 gradi di libertà e il criterio di informazione di Akaike è 41.43.

Si considerino ora due modifiche alla funzione logit, una alla coda destra e una per la coda sinistra prese separatamente. Le modifiche considerate sono le seguenti:

$$h_{\eta_0}(\eta, \psi = \psi_1) = \begin{cases} \eta_0 + \log(\eta - \eta_0 + 1) & \text{se } \eta \geq \eta_0 \text{ e } \psi_1 = 0 \\ \eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1} & \text{se } \eta \geq \eta_0 \text{ e } \psi_1 \neq 0 \\ \eta & \text{se } \eta < \eta_0, \end{cases} \quad (3.19)$$

$$h_{\eta_0}(\eta, \psi = \psi_2) = \begin{cases} \eta & \text{se } \eta \geq \eta_0 \\ \eta_0 - \log(-\eta + \eta_0 + 1) & \text{se } \eta < \eta_0 \text{ e } \psi_2 = 0 \\ \eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2} & \text{se } \eta < \eta_0 \text{ e } \psi_2 \neq 0. \end{cases} \quad (3.20)$$

Per studiare se l'introduzione dei parametri  $\psi_1$  e  $\psi_2$  porta o meno ad un miglioramento nell'adattamento del modello ai dati si considerano i grafici della devianza residua al variare dei parametri di legame.

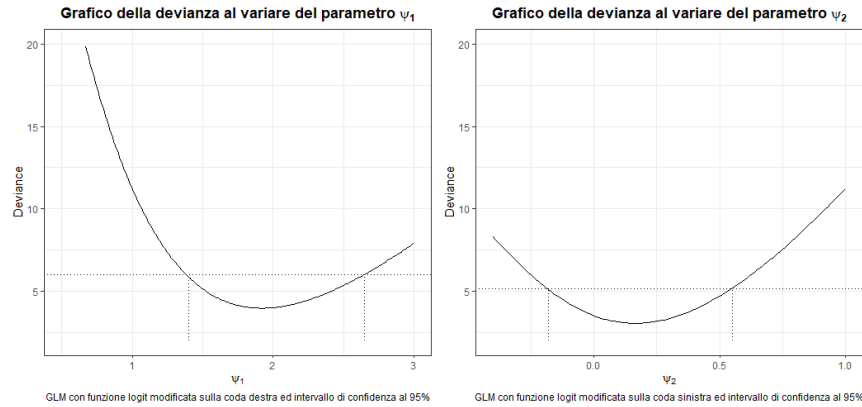


Figura 3.3.3: Devianza per il modello di regressione logistica per vari valori di  $\psi_1$ , a destra, e di  $\psi_2$  a sinistra, con  $\eta_0 = 0$ .

Le analisi effettuate suggeriscono un valore di  $\psi_1$  pari a 1.92 (e relativo intervallo di confidenza al 95% pari a (1.40, 2.65)) per la modifica della coda destra, e di  $\psi_2$  pari a 0.16 per quella alla coda sinistra (con intervallo al 95% pari a (-0.18, 0.55)). I valori che si ottengono per la devianza residua sono rispettivamente 3.96 e 3.05 su 6 gradi di libertà, mentre l'indice di Akaike è pari a 34.16 e 33.24.

Gli intervalli di confidenza al 95% per  $\psi_1$  e  $\psi_2$  non contengono il valore  $\psi = 1$ , allora le modifiche alle code destra e sinistra prese separatamente sono entrambe appropriate.

I risultati dell'adattamento dei due modelli sono riportati nelle Tabelle 3.3.5 e 3.3.4. In entrambi i casi la trasformazione che viene applicata porta ad un sostanziale miglioramento rispetto al modello di regressione logistica iniziale (Figura 3.3.4). Il miglioramento in termini di devianza che si ottiene, infatti, è pari a 64.75% e 72.89%. Tuttavia l'adattamento del modello con modifica alla coda destra è meno soddisfacente di quello che si ottiene apportando una modifica alla coda sinistra: sia l'indice di Akaike che la devianza residua sono infatti minori nel secondo caso.

Tabella 3.3.4: Output del modello di regressione logistica con coda destra modificata per il dataset *Beetles* con  $\psi_1 = 1.92$  e  $\eta_0 = 0$ .

Coefficienti	Stima	Std. Error	z value	p-value
Intercetta	-46.35	3.84	-12.06	< 0.001
dose	26.05	2.14	12.19	< 0.001

Tabella 3.3.5: Output del modello di regressione logistica con coda sinistra modificata per il dataset *Beetles* con  $\psi_2 = 0.16$  e  $\eta_0 = 0$ .

Coefficienti	Stima	Std. Error	z value	p-value
Intercetta	-86.39	9.81	-8.81	< 0.001
dose	48.45	5.46	8.88	< 0.001

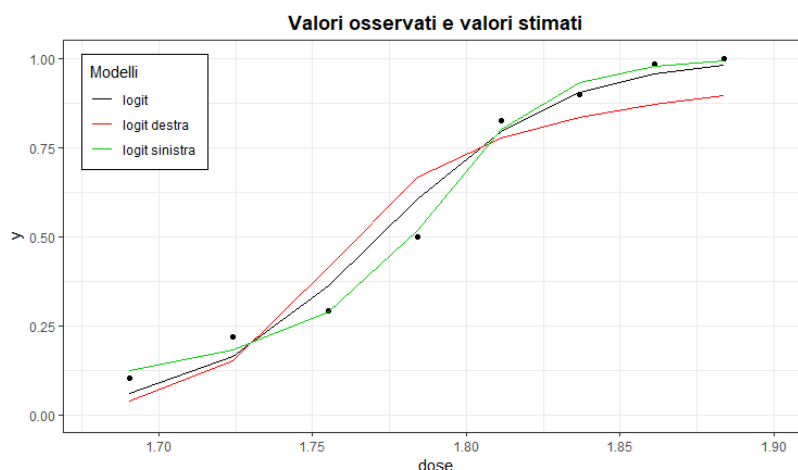


Figura 3.3.4: Modello logistico e modelli modificati con  $\psi_1 = 1.92$  e  $\psi_2 = 0.16$  e con  $\eta_0 = 0$ .

### 3.4 Conclusioni

In questo capitolo è stata presentata una tecnica di stima per i parametri di legame e di regressione attraverso la verosimiglianza profilo.

È stata posta l'attenzione sul beneficio che i parametri di legame apportano al modello in termini di bontà d'adattamento.

Ricordando che la funzione di legame canonica consente di ottenere risultati che sono facilmente interpretabili (si pensi ad esempio all'interpretazione del logaritmo della quota per il modello logistico), lo scopo è quello di ricercare il miglior modello che sappia coniugare tale aspetto con la bontà d'adattamento. Il procedimento ottimo, quindi, consiste nel confrontare sempre il

---

modello modificato con il modello di partenza (canonico).

Nelle analisi riportate si è mostrato un guadagno positivo in termini di devianza, ad indicare l'effettiva utilità del procedimento di selezione del miglior modello.

# Conclusioni

I GLM sono modelli essenziali per l'analisi statistica e per la trattazione di applicazioni pratiche a dati reali. La teoria sulla quale si basano è molto ampia e dà luogo a molte estensioni.

Un aspetto importante riguarda la specificazione stessa del modello: è fondamentale, infatti, indicare correttamente la funzione che lega la componente aleatoria a quella sistematica. Per la definizione di tali funzioni si ricorre a famiglie generatrici che permettono la costruzione di link per diverse distribuzioni.

Fondamentale è riuscire a trovare un bilanciamento tra l'aumento delle varianze nelle stime dei coefficienti di regressione ed il miglioramento nell'adattamento del modello ai dati attraverso l'introduzione dei parametri di link.

Si ricorre inoltre all'utilizzo di vari test d'ipotesi al fine di assicurarsi la miglior specificazione della funzione di legame per il problema in questione. Un secondo approccio per verificare, ad esempio, se l'utilizzo di un legame diverso da quello canonico è effettivamente utile ad apportare un miglioramento nell'adattamento ai dati, consiste nell'utilizzo dei fattori di Bayes. Indicando con  $\beta_\psi$  i parametri di regressione per un modello  $M_\psi$ , il fattore di Bayes per  $M_\psi$  contro il modello  $M_c$  a partire dai dati osservati  $Y = (Y_1, \dots, Y_n)$  è definito come

$$B_\psi = \frac{pr(Y|M_\psi)}{pr(Y|M_c)}. \quad (3.21)$$

dove

$$pr(Y|M_\psi) = \int pr(Y|M_\psi, \beta_\psi) pr(\beta_\psi|M_\psi) d\beta_\psi, \quad (3.22)$$

con  $\beta_\psi$  il parametro di regressione per il modello  $M_\psi$  e  $pr(\beta_\psi|M_\psi)$  la sua densità a priori nel modello  $M_\psi$ . Si possono confrontare  $B_{\psi_1}$ ,  $B_{\psi_2}$  e  $B_{\psi_1, \psi_2}$  per valori differenti del parametro di legame, oppure si possono costruire i fattori di Bayes per ogni modifica delle code.

La possibilità di poter estendere la funzione di legame attraverso delle trasformazioni di potenza, quando è realmente necessario, consente di ottenere una maggiore flessibilità per i GLM, portando ad una diminuzione della

devianza residua.

A seconda del caso di studio può esser utile riuscire a governare il comportamento delle code e dunque la simmetria della funzione di legame.

Viene riportato un esempio (Figura 3.4.1) della modifica simultanea di entrambe le code per una funzione di legame probit al variare dei due parametri di legame  $\psi_1$  e  $\psi_2$ .

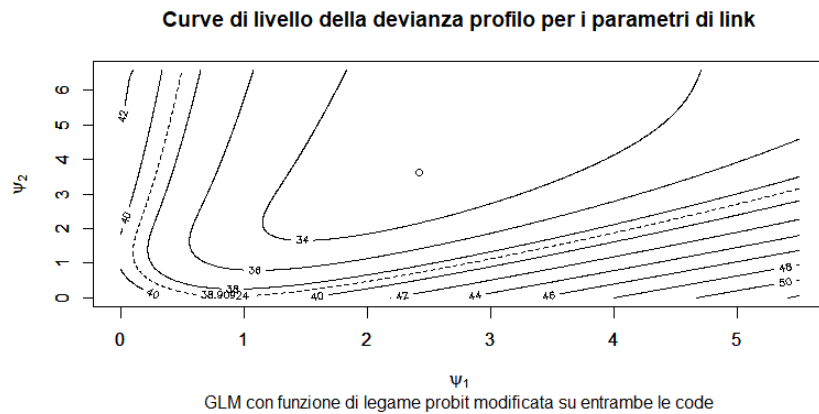


Figura 3.4.1: Curve di livello per la devianza del modello di regressione con funzione di legame probit per vari valori di  $\psi_1$  e  $\psi_2$  con  $\eta_0 = 0$ .

La stima dei parametri di legame può comportare difficoltà dal punto di vista computazionale. Un modo di procedere, allora, può essere guardare come varia la devianza residua al variare di una griglia di valori per il parametro di link. In questo esempio, le curve di livello della devianza per vari valori del parametro  $\psi = (\psi_1, \psi_2)$ , portano ad individuare una stima per  $\psi$ .

# Bibliografia

- Agresti A. (2015). *Foundations of Linear and Generalized Linear Models*. Wiley.
- Aitkin M.(2018). *A History of the GLIM Statistical Package*. International Statistical Review. **86**(2), 275–299.
- Aranda-Ordaz F. J.(1981). *On two families of transformations to Additivity for Binary Response data*. Biometrika. **68**(2), 357–363.
- Bellio R.; Pierce D.(2018). *Functions for Likelihood Asymptotics*.
- Bertsch M.; Dal Passo R.; Giacomelli L. (2007). *Analisi matematica*. McGraw-Hill.
- Bliss C. I.(1935). *The Calculation of the Dosage-Mortality Curve*. Annals of Applied Biology. **22**(1), 134–167.
- Bortot P.; Ventura L.; Salvan A. (2000). *Inferenza Statistica: Applicazioni con S-PLUS e R*. CEDAM.
- Brazzale A. R.; Davison A. C.; Reid N. (2007). *Applied Asymptotics Case Studies in Small-Sample Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics.
- Brillinger D.; Fienberg S.; Gani J.; Hartigan J.; Krickeberg K.; Gilchrist R. (1985). *Lecture Notes in Statistics Generalized Linear Models*.
- Burr I. W.(1942). *Cumulative Frequency Functions*. Ann. Math. Statist.. **2**, 215–232.
- Cox D. R.; Reid N.(2013). *Parameter Orthogonality and Approximate Conditional Inference*. Journal of the Royal Statistical Society B. **49**(1), 1–39.
- Czado C.(1994). *Parametric link modification of both tails in binary regression*. Statistical Papers. **35**, 189–201.

- Czado C.(1997). *On Selecting Parametric Link Transformation Families in Generalized Linear Models*. Journal of Statistical Planning and Inference. **61**(1).
- Czado C.; Munk A.(2000). *Noncanonical links in generalized linear models - when is the effort justified?* Journal of Statistical Planning and Inference. **87**(2), 317–345.
- Czado C.; Santner T. J.(1992). *Orthogonalizing parametric link transformation families in binary regression analysis*. Canadian Journal of Statistic. **20**(I), 51–61.
- Fahrmeir L.; Francis B.; Gilchrist R.; Tutz G. (1992). *Advances in GLIM and Statistical Modelling*. Springer-Verlag.
- Flach N.(2014). *Generalized linear models with parametric link families in R*. Thesis.
- Grigoletto M.; Pauli F.; Ventura L. (2017). *Modello lineare Teoria e applicazione con R*. G. Giappichelli.
- McCullagh P.; Nelder J. A. (1989). *Generalized Linear Models*. Chapman and Hall. Boston, MA.
- Nelder J.; Wedderburn R.(1972). *Generalized Linear Models*. Journal of the Royal Statistical Society. **135**(3), 370–384.
- Pace L.; Salvan A. (2001). *Introduzione alla statistica - Inferenza, verosimiglianza, modelli*. CEDAM. Padova.
- Pregibon D.(1980). *Goodness of Link Tests for Generalized Linear Models*. Applied Statistics. **29**(1), 15–23.
- Prentice R. L.(1976). *Generalization of the probit and Logit Methods for Dose Response Curves*. Biometrics. **32**(4), 761–768.
- Raftery A.(1996). *Approximate Bayes factors and accounting for model uncertainty in generalised linear models*. Biometrika. **83**(2), 251–266.
- Stukel T. A.(1989). *Generalized logistic models*. Mathematical and Computer Modelling. **12**(9), 1182.
- Ventura L.; Racugno W. (2017). *Biostatistica. Casi di studio in R*. Egea. Milano.



# Appendice

## Codice R

Si riporta il codice R utilizzato per l'implementazione degli esempi.

```
# Famiglia parametrica di Stukel

hstu2 <- function(alpha1=stop("Argument 'alpha1' is missing"),
                  alpha2=stop("Argument 'alpha2' is missing"),
                  eta=stop("Argument 'eta' is missing")){
  h = 1:length(eta)
  if(alpha1>0){
    h[eta>=0] = alpha1^(-1)*(exp(alpha1*abs(eta[eta>=0]))-1)
  }
  else if(alpha1==0){
    h[eta>=0] = eta[eta>=0]
  }
  else{
    h[eta>=0] = -alpha1^(-1)*log(1-alpha1*abs(eta[eta>0]))
  }
  if(alpha2>0){
    h[eta<0] = -alpha2^(-1)*(exp(alpha2*abs(eta[eta<0]))-1)
  }
  else if(alpha2==0){
    h[eta<0]=eta[eta<0]
  }
  else{
    h[eta<0]=alpha2^(-1)*log(1-alpha2*abs(eta[eta<0]))
  }
  h
}

# Esempio con libreria likelihoodAsy

library(likelihoodAsy)
```

```
mod.glm <- glm(formula = cbind(r1, r2) ~ day + lull - 1,
               family = binomial, data = babies)
data.obj <- list(y = babies$r1, den = babies$r1 + babies$r2,
               X = model.matrix(mod.glm))

# Log-verosimiglianza con funzione di legame parametrica
# di Burr-Prentice

loglik.logit.pb<- function(theta, data)
{
  y <- data$y
  den <- data$den
  X <- data$X
  eta <- X %*% theta[1:19]
  psi=theta[20]
  pb = 1-(1/(1+exp(eta))^psi)
  l <- sum(y * log(pb) + (den - y) * log(1-pb))
  return(l)
}

# Funzione generatrice dei dati

gendat.logit.pb<- function(theta, data)
{
  X <- data$X
  eta <- X %*% theta[1:19]
  psi=theta[20]
  pb = 1-(1/(1+exp(eta))^psi)
  out <- data
  out$y <- rbinom(length(data$y), size = data$den, prob = pb)
  return(out)
}

# Calcolo rstar

rstar(data=data.obj, thetainit = c(coef(mod.glm),1),
      floglik = loglik.logit.pb, fpsi = function(theta) theta[20],
      fscore=NULL, datagen=gendat.logit.pb, trace=FALSE, psival=1)
```